

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

GenCore version 5.1.4.p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:30:32 ; Search time 8.35496 Seconds
(without alignments)
1097.829 Million cell updates/sec

Title: US-09-816-653A-2

Perfect score: 1061
Sequence: 1 MSQPNQNTSPGATPDYSG.....MAHLGVQVMORACDPQVPI 199

Scoring table: BLASTN62
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEM_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PC1_NEM_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEM_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEM_PUB pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEM_PUB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEM_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1059	99.8	199	US-09-816-653A-2	Sequence 2, Appli
2	1048	98.8	667	US-09-759-056-2	Sequence 2, Appli
3	1048	98.8	667	US-09-901-812-2	Sequence 2, Appli
4	1048	98.8	667	US-10-227-884-80	Sequence 80, Appli
5	1048	98.8	667	US-10-230-163-80	Sequence 80, Appli
6	1048	98.8	667	US-10-218-631-80	Sequence 80, Appli
7	1048	98.8	667	US-10-230-338-80	Sequence 80, Appli
8	986.5	93.0	658	US-09-759-056-5	Sequence 5, Appli
9	986.5	93.0	658	US-09-901-812-5	Sequence 5, Appli
10	697	65.7	670	US-09-816-653A-7	Sequence 7, Appli
11	84	7.9	624	US-09-925-297-738	Sequence 738, App
12	83	7.8	913	US-09-223-490-4	Sequence 4, Appli
13	78.5	7.4	3519	US-09-808-880-4	Sequence 4, Appli
14	74.5	7.0	208	US-09-738-626-6358	Sequence 6358, Ap
15	74	7.0	326	US-09-925-301-1349	Sequence 1349, Ap
16	73.5	6.9	356	US-09-738-626-4781	Sequence 4781, Ap
17	73.5	6.9	378	US-10-021-758-2	Sequence 2, Appli
18	73.5	6.9	636	US-10-000-776-12	Sequence 12, Appli
19	73.5	6.9	636	US-09-791-497-14	Sequence 14, Appli

20	73	6.9	633	9	US-10-066-464-11	Sequence 11, Appli
21	72.5	6.8	504	10	US-09-758-386-2	Sequence 2, Appli
22	72.5	6.8	504	10	US-09-815-108-17	Sequence 17, Appli
23	72.5	6.8	504	10	US-09-815-108-19	Sequence 19, Appli
24	72.5	6.8	529	9	US-09-712-363-175	Sequence 175, App
25	72	6.8	183	10	US-09-764-853-822	Sequence 822, App
26	72	6.8	652	10	US-10-139-662-4	Sequence 4, Appli
27	72	6.8	694	10	US-09-846-996A-1	Sequence 1, Appli
28	70.5	6.6	627	10	US-09-853-386-67	Sequence 67, Appli
29	70.5	6.6	627	12	US-10-005-169-2	Sequence 2, Appli
30	70.5	6.6	628	10	US-09-853-386-66	Sequence 66, Appli
31	70.5	6.6	628	10	US-09-853-386-68	Sequence 68, Appli
32	70.5	6.6	628	10	US-09-853-386-73	Sequence 73, Appli
33	70.5	6.6	628	12	US-10-005-169-4	Sequence 4, Appli
34	70.5	6.6	730	9	US-09-738-626-4196	Sequence 4196, Ap
35	69.5	6.6	206	10	US-09-800-729-149	Sequence 149, App
36	69.5	6.6	369	12	US-10-077-874-4	Sequence 4, Appli
37	69.5	6.6	372	9	US-09-393-696-6	Sequence 6, Appli
38	69.5	6.6	372	12	US-10-077-874-6	Sequence 6, Appli
39	69.5	6.6	389	10	US-09-828-538-20	Sequence 20, Appli
40	69.5	6.6	425	10	US-09-828-538-19	Sequence 19, Appli
41	69.5	6.6	425	10	US-09-828-538-24	Sequence 24, Appli
42	69.5	6.6	650	9	US-10-139-662-2	Sequence 2, Appli
43	69	6.5	122	10	US-09-796-858-14	Sequence 14, Appli
44	69	6.5	238	9	US-10-001-054-34	Sequence 34, Appli
45	69	6.5	238	9	US-10-006-856A-271	Sequence 271, App

ALIGNMENTS

RESULT 1
US-09-816-653A-2
Sequence 2, Application US/0981653A
Publication No. US20030021788A1
GENERAL INFORMATION:
APPLICANT: Rastelli, Luca
TITLE OR INVENTION: NOVEL HUMAN STRA6-LIKE PROTEIN AND NUCLEIC ACIDS ENCODING THE SAME
FILE REFERENCE: 10716/57
CURRENT APPLICATION NUMBER: US/09/816,653A
CURRENT FILING DATE: 2001-03-23
PRIORITY APPLICATION NUMBER: 60/191,532
PRIOR FILING DATE: 2000-03-23
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 199
TYPE: PRT
ORGANISM: Homo sapiens hSTRA6 amino terminal polypeptide fragment
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (31)..(31)
OTHER INFORMATION: Xaa is any amino acid
US-09-816-653A-2

Query Match 99.8%; Score 1059; DB 9; Length 199;
Best Local Similarity 100.0%; Pred. No. 1.3e-95;
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSQPNQNTSPGATPDYSGWYIDEPQGXELQEGEVPSCHTSIIPGLYHACIASLSI 60
DB 1 MSQPNQNTSPGATPDYSGWYIDEPQGXELQEGEVPSCHTSIIPGLYHACIASLSI 60
QY 61 LVLLILAMLVRRRQLMPDCVGRPGLPSPVDFLAGDRPAVPAVFWVLLSLCLLPDE 120
DB 61 LVLLILAMLVRRRQLMPDCVGRPGLPSPVDFLAGDRPAVPAVFWVLLSLCLLPDE 120
QY 121 DALPFTTILASASQKTEAPRGAKIIGLFYALYPLAACATAGHTAHLIGSTLSW 180
DB 121 DALPFTTILASASQKTEAPRGAKIIGLFYALYPLAACATAGHTAHLIGSTLSW 180
QY 181 AHLGVQVMORACDPQVPI 199

Db : 181 AHGVQVWQRAECPOVPKI 199

RESULT 2

US-09-759-056-2
Sequence 2, Application US/09759056
Patent No. US20020156252A1
GENERAL INFORMATION:
APPLICANT: Pennica, Diane
APPLICANT: Smith, Victoria
TITLE OF INVENTION: No. US20020156252A1 STRA6 Polypeptides
FILE REFERENCE: GENENT.2827A2
CURRENT APPLICATION NUMBER: US/09/759,056
PRIOR FILING DATE: 2001-01-11
PRIOR APPLICATION NUMBER: 60/197089
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/175849
PRIOR FILING DATE: 2000-01-13
PRIOR APPLICATION NUMBER: 60/228914
PRIOR FILING DATE: 2000-08-29
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 667
TYPE: PRT
ORGANISM: Homo Sapiens
US-09-759-056-2

Query Match 98.8%; Score 1048; DB 9; Length 667;
Best Local Similarity 99.0%; Pred. No. 6,5e-94;

Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SOPAGNOTSGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSPGLYHACLASLSTL 61
3 SOPAGNOTSGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSPGLYHACLASLSTL 62
QY 62 VLLLLAMLVRRQQLWPDVCVRGRLPSVDFLAGDRPRAVPAVFMVLLSSCLLLPDED 121
63 VLLLLAMLVRRQQLWPDVCVRGRLPSVDFLAGDRPRAVPAVFMVLLSSCLLLPDED 122
QY 122 ALPFLTLASPSQDGTEAPRGWKILGLFYAALYYPLAACATAGHTAAHLGSLTSLWA 181
123 ALPFLTLASPSQDGTEAPRGWKILGLFYAALYYPLAACATAGHTAAHLGSLTSLWA 182
QY 182 HLGQVQWQRAECPOVPKI 199
183 HLGQVQWQRAECPOVPKI 200

RESULT 3

US-09-901-812-2
Sequence 2, Application US/09901812
Patent No. US20020173461A1
GENERAL INFORMATION:
APPLICANT: Pennica, Diane
APPLICANT: Polak, Paul
TITLE OF INVENTION: UPREGULATION OF TUMOR ANTIGENS TO
FILE REFERENCE: GENENT.081A
CURRENT APPLICATION NUMBER: US/09/901,812
PRIOR FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/228,914
PRIOR FILING DATE: 2000-08-29
PRIOR APPLICATION NUMBER: 09/759,056
PRIOR FILING DATE: 2001-01-11
PRIOR APPLICATION NUMBER: 60/175,849
PRIOR FILING DATE: 2000-01-13
PRIOR APPLICATION NUMBER: 60/197,089
NUMBER OF SEQ ID NOS: 28

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 667
TYPE: PRT
ORGANISM: Homo sapiens
US-09-901-812-2

Query Match

Best Local Similarity 98.8%; Score 1048; DB 9; Length 667;
Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SOPAGNOTSGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSPGLYHACLASLSTL 61
3 SOPAGNOTSGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSPGLYHACLASLSTL 62
QY 62 VLLLLAMLVRRQQLWPDVCVRGRLPSVDFLAGDRPRAVPAVFMVLLSSCLLLPDED 121
63 VLLLLAMLVRRQQLWPDVCVRGRLPSVDFLAGDRPRAVPAVFMVLLSSCLLLPDED 122
QY 122 ALPFLTLASPSQDGTEAPRGWKILGLFYAALYYPLAACATAGHTAAHLGSLTSLWA 181
123 ALPFLTLASPSQDGTEAPRGWKILGLFYAALYYPLAACATAGHTAAHLGSLTSLWA 182
QY 182 HLGQVQWQRAECPOVPKI 199
183 HLGQVQWQRAECPOVPKI 200

RESULT 4

US-10-227-884-80
Sequence 80, Application US/10227884
Publication No. US20030027988A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Desnoyers, Luc
APPLICANT: Geritsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Watanabe, Colin L.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P350P1C79
CURRENT APPLICATION NUMBER: US/10/227,884
PRIOR FILING DATE: 2002-08-26
PRIOR APPLICATION NUMBER: 10/119,480
PRIOR FILING DATE: 2002-04-09
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063549
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/069873
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15

PRIOR APPLICATION NUMBER: 60/082804
 PRIOR FILING DATE: 1998-04-22
 PRIOR APPLICATION NUMBER: 60/084441
 PRIOR FILING DATE: 1998-05-06
 PRIOR APPLICATION NUMBER: 60/085323
 PRIOR FILING DATE: 1998-05-13
 PRIOR APPLICATION NUMBER: 60/085579
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 60/086392
 PRIOR FILING DATE: 1998-05-22
 PRIOR APPLICATION NUMBER: 60/089532
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089538
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089905
 PRIOR FILING DATE: 1998-06-18
 PRIOR APPLICATION NUMBER: 60/090472
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090557
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090691
 PRIOR FILING DATE: 1998-06-25
 PRIOR APPLICATION NUMBER: 60/090695
 PRIOR FILING DATE: 1998-06-25
 PRIOR APPLICATION NUMBER: 60/091982
 PRIOR FILING DATE: 1998-07-07
 PRIOR APPLICATION NUMBER: 60/095302
 PRIOR FILING DATE: 1998-08-04
 PRIOR APPLICATION NUMBER: 60/095318
 PRIOR FILING DATE: 1998-08-04
 PRIOR APPLICATION NUMBER: 60/095916
 PRIOR FILING DATE: 1998-08-10
 PRIOR APPLICATION NUMBER: 60/096146
 PRIOR FILING DATE: 1998-08-11
 PRIOR APPLICATION NUMBER: 60/096791
 PRIOR FILING DATE: 1998-08-17
 PRIOR APPLICATION NUMBER: 60/097986
 PRIOR FILING DATE: 1998-08-26
 PRIOR APPLICATION NUMBER: 60/098544
 PRIOR FILING DATE: 1998-08-31
 PRIOR APPLICATION NUMBER: 60/099596
 PRIOR FILING DATE: 1998-09-09
 PRIOR APPLICATION NUMBER: 60/099598
 PRIOR FILING DATE: 1998-09-09
 PRIOR APPLICATION NUMBER: 60/099803
 PRIOR FILING DATE: 1998-09-10
 PRIOR APPLICATION NUMBER: 60/099811
 PRIOR FILING DATE: 1998-09-10
 PRIOR APPLICATION NUMBER: 60/099812
 PRIOR FILING DATE: 1998-09-10
 PRIOR APPLICATION NUMBER: 60/099816
 PRIOR FILING DATE: 1998-09-10
 PRIOR APPLICATION NUMBER: 60/100038
 PRIOR FILING DATE: 1998-09-11
 PRIOR APPLICATION NUMBER: 60/100385
 PRIOR FILING DATE: 1998-09-15
 PRIOR APPLICATION NUMBER: 60/100390
 PRIOR FILING DATE: 1998-09-15
 PRIOR APPLICATION NUMBER: 60/100627
 PRIOR FILING DATE: 1998-09-16
 PRIOR APPLICATION NUMBER: 60/100848
 PRIOR FILING DATE: 1998-09-18
 PRIOR APPLICATION NUMBER: 60/100919
 PRIOR FILING DATE: 1998-09-17
 PRIOR APPLICATION NUMBER: 60/101477
 PRIOR FILING DATE: 1998-09-23
 PRIOR APPLICATION NUMBER: 60/101738
 PRIOR FILING DATE: 1998-09-24
 PRIOR APPLICATION NUMBER: 60/101741
 PRIOR FILING DATE: 1998-09-24
 PRIOR APPLICATION NUMBER: 60/101786
 PRIOR FILING DATE: 1998-09-25
 PRIOR APPLICATION NUMBER: 60/101916

PRIOR FILING DATE: 1998-09-24
 PRIOR APPLICATION NUMBER: 60/101922
 PRIOR FILING DATE: 1998-09-24
 PRIOR APPLICATION NUMBER: 60/106178
 PRIOR FILING DATE: 1998-10-28
 PRIOR APPLICATION NUMBER: 60/106248
 PRIOR FILING DATE: 1998-10-29
 PRIOR APPLICATION NUMBER: 60/106464
 PRIOR FILING DATE: 1998-10-30
 PRIOR APPLICATION NUMBER: 60/106905
 PRIOR FILING DATE: 1998-11-03
 PRIOR APPLICATION NUMBER: 60/108787
 PRIOR FILING DATE: 1998-11-17
 PRIOR APPLICATION NUMBER: 60/108801
 PRIOR FILING DATE: 1998-11-17
 PRIOR APPLICATION NUMBER: 60/108849
 PRIOR FILING DATE: 1998-11-18
 PRIOR APPLICATION NUMBER: 60/112422
 PRIOR FILING DATE: 1998-12-15
 PRIOR APPLICATION NUMBER: 60/113296
 PRIOR FILING DATE: 1998-12-22
 PRIOR APPLICATION NUMBER: 60/113605
 PRIOR FILING DATE: 1998-12-23
 PRIOR APPLICATION NUMBER: 60/113621
 PRIOR FILING DATE: 1998-12-23
 PRIOR APPLICATION NUMBER: 60/115558
 PRIOR FILING DATE: 1999-01-12
 PRIOR APPLICATION NUMBER: 60/115565
 PRIOR FILING DATE: 1999-01-12
 PRIOR APPLICATION NUMBER: 60/115733
 PRIOR FILING DATE: 1999-01-12
 PRIOR APPLICATION NUMBER: 60/119549
 PRIOR FILING DATE: 1999-02-10
 PRIOR APPLICATION NUMBER: 60/123618
 PRIOR FILING DATE: 1999-03-10
 PRIOR APPLICATION NUMBER: 60/125259
 PRIOR FILING DATE: 1999-03-19
 PRIOR APPLICATION NUMBER: 60/125775
 PRIOR FILING DATE: 1999-03-23
 PRIOR APPLICATION NUMBER: 60/126773
 PRIOR FILING DATE: 1999-03-29
 PRIOR APPLICATION NUMBER: 60/127887
 PRIOR FILING DATE: 1999-04-05
 PRIOR APPLICATION NUMBER: 60/130232
 PRIOR FILING DATE: 1999-04-21
 PRIOR APPLICATION NUMBER: 60/131022
 PRIOR FILING DATE: 1999-04-26
 PRIOR APPLICATION NUMBER: 60/131270
 PRIOR FILING DATE: 1999-04-27
 PRIOR APPLICATION NUMBER: 60/131291
 PRIOR FILING DATE: 1999-04-27
 PRIOR APPLICATION NUMBER: 60/131445
 PRIOR FILING DATE: 1999-04-28
 PRIOR APPLICATION NUMBER: 60/134287
 PRIOR FILING DATE: 1999-05-14
 PRIOR APPLICATION NUMBER: 60/140650
 PRIOR FILING DATE: 1999-06-22
 PRIOR APPLICATION NUMBER: 60/140723
 PRIOR FILING DATE: 1999-06-22
 PRIOR APPLICATION NUMBER: 60/141037
 PRIOR FILING DATE: 1999-06-23
 PRIOR APPLICATION NUMBER: 60/144758
 PRIOR FILING DATE: 1999-07-20
 PRIOR APPLICATION NUMBER: 60/145698
 PRIOR FILING DATE: 1999-07-26
 PRIOR APPLICATION NUMBER: 60/146222
 PRIOR FILING DATE: 1999-07-28
 PRIOR APPLICATION NUMBER: 60/146963
 PRIOR FILING DATE: 1999-08-03
 PRIOR APPLICATION NUMBER: 60/149320
 PRIOR FILING DATE: 1999-08-17
 PRIOR APPLICATION NUMBER: 60/149638
 PRIOR FILING DATE: 1999-08-17

PRIOR APPLICATION NUMBER: 60/151733
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/164418
PRIOR FILING DATE: 1999-11-09
PRIOR APPLICATION NUMBER: 60/166361
PRIOR FILING DATE: 1999-11-16
PRIOR APPLICATION NUMBER: 60/169445
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: 60/169495
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: 60/169835

Query Match 98.8%; Score 1048; DB 9; Length 667;
Best Local Similarity 99.0%; Pred. No. 6.5e-94;
Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SOPAGNOSTGATEDYSGWYIDEPGGKELDPEGEVSCHTSIPGLYHACLASLIL 61
DB 3 SOPAGNOSTGATEDYSGWYIDEPGGKELDPEGEVSCHTSIPGLYHACLASLIL 62
QY 62 VLLLLMLVRRRLMPDCVGRGLSPVDPLAGDRPRAVPAVFMVLLSLCLLLPDED 121
DB 63 VLLLLMLVRRRLMPDCVGRGLSPVDPLAGDRPRAVPAVFMVLLSLCLLLPDED 122
QY 122 ALPFLTLASPSODGKTEAPRGAMKILGLFHYALYPLAACATAGHTAHLGLSTLSWA 181
DB 123 ALPFLTLASPSODGKTEAPRGAMKILGLFHYALYPLAACATAGHTAHLGLSTLSWA 182
QY 182 HLGVOVMORAECPQVPI 199
DB 183 HLGVOVMORAECPQVPI 200

RESULT 5
US-10-230-163-80
Sequence 80, Application US/10230163
Publication No. US20030036635A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Desnoyers, Luc
APPLICANT: Gerltzen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Goddard, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Guirney, Austin L.
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Macanabe, Colin L.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3530P1C96
CURRENT APPLICATION NUMBER: US/10/230,163
CURRENT FILING DATE: 2002-08-28
PRIOR APPLICATION NUMBER: 10/119,480
PRIOR FILING DATE: 2002-04-09
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063549
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/069873
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079728

PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089905
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090691
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/095302
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/095318
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/095916
PRIOR FILING DATE: 1998-08-10
PRIOR APPLICATION NUMBER: 60/096146
PRIOR FILING DATE: 1998-08-11
PRIOR APPLICATION NUMBER: 60/096791
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: 60/097986
PRIOR FILING DATE: 1998-08-26
PRIOR APPLICATION NUMBER: 60/098544
PRIOR FILING DATE: 1998-08-31
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099803
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099811
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100038
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100848
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24

PRIOR APPLICATION NUMBER: 60/101741
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101766
PRIOR FILING DATE: 1998-09-25
PRIOR APPLICATION NUMBER: 60/101916
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101922
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/106178
PRIOR FILING DATE: 1998-10-28
PRIOR APPLICATION NUMBER: 60/106248
PRIOR FILING DATE: 1998-10-29
PRIOR APPLICATION NUMBER: 60/106464
PRIOR FILING DATE: 1998-10-30
PRIOR APPLICATION NUMBER: 60/106905
PRIOR FILING DATE: 1998-11-03
PRIOR APPLICATION NUMBER: 60/108787
PRIOR FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: 60/108801
PRIOR FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: 60/108849
PRIOR FILING DATE: 1998-11-18
PRIOR APPLICATION NUMBER: 60/112422
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: 60/113296
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/113605
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113621
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/115558
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115565
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115733
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/119549
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/123618
PRIOR FILING DATE: 1999-03-10
PRIOR APPLICATION NUMBER: 60/125259
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: 60/125775
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/126773
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: 60/127887
PRIOR FILING DATE: 1999-04-05
PRIOR APPLICATION NUMBER: 60/130232
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/131022
PRIOR FILING DATE: 1999-04-26
PRIOR APPLICATION NUMBER: 60/131270
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131291
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131445
PRIOR FILING DATE: 1999-04-28
PRIOR APPLICATION NUMBER: 60/134287
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/140650
PRIOR FILING DATE: 1999-06-22
PRIOR APPLICATION NUMBER: 60/140723
PRIOR FILING DATE: 1999-06-22
PRIOR APPLICATION NUMBER: 60/141037
PRIOR FILING DATE: 1999-06-23
PRIOR APPLICATION NUMBER: 60/144758
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/145698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 60/146222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: 60/146963

PRIOR FILING DATE: 1999-08-03
PRIOR APPLICATION NUMBER: 60/149320
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/149638
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/151733
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/164418
PRIOR FILING DATE: 1999-11-09
PRIOR APPLICATION NUMBER: 60/166361
PRIOR FILING DATE: 1999-11-16
PRIOR APPLICATION NUMBER: 60/169445
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: 60/169495
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: 60/169835

Query Match 98.8%; Score 1048; DB 9; Length 667;
Best Local Similarity 99.0%; Pred. No. 6, 5e-94;
Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SOPAGNOTSGATEDSYSGWYIDEPQGXELQPEGEVPSCHTSPGLYHACLASLIL 61
DB 3 SOPAGNOTSGATEDSYSGWYIDEPQGEELQPEGEVPSCHTSPGLYHACLASLIL 62
QY 62 VLLLLAMLVRRQLMPDCVGRGRLPSVDFLAGDRPRAVPAVFNVLSSICLLIPDED 121
DB 63 VLLLLAMLVRRQLMPDCVGRGRLPSVDFLAGDRPRAVPAVFNVLSSICLLIPDED 122
QY 122 ALPPLTLASAPSDGKTEARAGMKILGLPHYAALVYPLAACATAGHTAAHLGSLTSLNA 181
DB 123 ALPPLTLASAPSDGKTEARAGMKILGLPHYAALVYPLAACATAGHTAAHLGSLTSLNA 182
QY 182 HLGQVQWQRAECQPVPKI 199
DB 183 HLGQVQWQRAECQPVPKI 200

RESULT 6
US-10-218-631-80
Sequence 80: Application US//10218631
Publication No. US20030045667A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Desnoyers, Luc
APPLICANT: Gerritsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Watanabe, Colin L.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3530P1C14
CURRENT APPLICATION NUMBER: US//10/218, 631
CURRENT FILING DATE: 2002-08-12
PRIOR APPLICATION NUMBER: 10/119, 480
PRIOR FILING DATE: 2002-04-09
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063549
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/069873
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20

;; PRIOR APPLICATION NUMBER: 60/079294
;; PRIOR FILING DATE: 1998-03-25
;; PRIOR APPLICATION NUMBER: 60/079656
;; PRIOR FILING DATE: 1998-03-26
;; PRIOR APPLICATION NUMBER: 60/079728
;; PRIOR FILING DATE: 1998-03-27
;; Remaining Prior Application data removed - See file Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 246
;; SEQ ID NO 80
;; LENGTH: 667
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-218-631-80

Query Match 98.8%; Score 1048; DB 9; Length 667;
Best Local Similarity 99.0%; Pred. No. 6,5e-94;
Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SOPAGNQTSPGATEDSYSGSWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLSIL 61
DB 3 SOPAGNQTSPGATEDSYSGSWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLSIL 62
QY 62 VLLLLAMLVRRQLMPDCVGRGRLPSVDFLAGDRPRAVPAVFMVLLSSLCILLPDED 121
DB 63 VLLLLAMLVRRQLMPDCVGRGRLPSVDFLAGDRPRAVPAVFMVLLSSLCILLPDED 122
QY 122 ALPFLTLASAPSDQKTEAPRGAMKILGLFYAALYYPPLAACATAGHTAAHLGSLTSLWA 181
DB 123 ALPFLTLASAPSDQKTEAPRGAMKILGLFYAALYYPPLAACATAGHTAAHLGSLTSLWA 182
QY 182 HLGVOVWQRAECPOVPKI 199
DB 183 HLGVOVWQRAECPOVPKI 200

RESULT 7

US-10-230-338-80
;; Sequence 80, Application US/10230338
;; Publication No. US20030044934A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin P.
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Gerlitsen, Mary
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gueney, Aubertin L.
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stephan, Jean-Philippe F.
;; APPLICANT: Watanabe, Colin L.
;; APPLICANT: Wood, William I.
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3530P1C92
;; CURRENT APPLICATION NUMBER: US/10/230,338
;; PRIOR FILING DATE: 2002-08-28
;; PRIOR APPLICATION NUMBER: 10/119,480
;; PRIOR FILING DATE: 2002-04-09
;; PRIOR APPLICATION NUMBER: 60/059113
;; PRIOR FILING DATE: 1997-09-17
;; PRIOR APPLICATION NUMBER: 60/062287
;; PRIOR FILING DATE: 1997-10-17
;; PRIOR APPLICATION NUMBER: 60/063549
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/064103
;; PRIOR FILING DATE: 1997-10-31
;; PRIOR APPLICATION NUMBER: 60/069873
;; PRIOR FILING DATE: 1997-12-17
;; PRIOR APPLICATION NUMBER: 60/078910
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/079294
;; PRIOR FILING DATE: 1998-03-25
;; PRIOR APPLICATION NUMBER: 60/079656

;; PRIOR FILING DATE: 1998-03-26
;; PRIOR APPLICATION NUMBER: 60/079728
;; PRIOR FILING DATE: 1998-03-27
;; Remaining Prior Application data removed - See file Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 246
;; SEQ ID NO 80
;; LENGTH: 667
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-230-338-80

Query Match 98.8%; Score 1048; DB 9; Length 667;
Best Local Similarity 99.0%; Pred. No. 6,5e-94;
Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SOPAGNQTSPGATEDSYSGSWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLSIL 61
DB 3 SOPAGNQTSPGATEDSYSGSWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLSIL 62
QY 62 VLLLLAMLVRRQLMPDCVGRGRLPSVDFLAGDRPRAVPAVFMVLLSSLCILLPDED 121
DB 63 VLLLLAMLVRRQLMPDCVGRGRLPSVDFLAGDRPRAVPAVFMVLLSSLCILLPDED 122
QY 122 ALPFLTLASAPSDQKTEAPRGAMKILGLFYAALYYPPLAACATAGHTAAHLGSLTSLWA 181
DB 123 ALPFLTLASAPSDQKTEAPRGAMKILGLFYAALYYPPLAACATAGHTAAHLGSLTSLWA 182
QY 182 HLGVOVWQRAECPOVPKI 199
DB 183 HLGVOVWQRAECPOVPKI 200

RESULT 8

US-09-759-056-5
;; Sequence 5, Application US/09759056
;; Patent No. US20020156252A1
;; GENERAL INFORMATION:
;; APPLICANT: Pennica, Diane
;; APPLICANT: Smith, Victoria
;; APPLICANT: Wood, William I.
;; TITLE OF INVENTION: NO. US20020156252A1e1 STRA6 Polypeptides
;; FILE REFERENCE: GENENT. 2827A2
;; CURRENT APPLICATION NUMBER: US/09/759,056
;; PRIOR FILING DATE: 2001-01-11
;; PRIOR APPLICATION NUMBER: 60/197089
;; PRIOR FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/175849
;; PRIOR FILING DATE: 2000-01-13
;; PRIOR APPLICATION NUMBER: 60/228914
;; PRIOR FILING DATE: 2000-08-29
;; NUMBER OF SEQ ID NOS: 5
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 5
;; LENGTH: 658
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-759-056-5

Query Match 93.0%; Score 986.5; DB 9; Length 658;
Best Local Similarity 94.4%; Pred. No. 6,1e-88;
Matches 187; Conservative 1; Mismatches 1; Indels 9; Gaps 1;

QY 2 SOPAGNQTSPGATEDSYSGSWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLSIL 61
DB 3 SOPAGNQTSPGATEDSYSGSWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLSIL 62
QY 62 VLLLLAMLVRRQLMPDCVGRGRLPSVDFLAGDRPRAVPAVFMVLLSSLCILLPDED 121
DB 63 VLLLLAMLVRRQLMPDCVGRGRLPSVDFLAGDRPRAVPAVFMVLLSSLCILLPDED 113
QY 122 ALPFLTLASAPSDQKTEAPRGAMKILGLFYAALYYPPLAACATAGHTAAHLGSLTSLWA 181
DB 114 ALPFLTLASAPSDQKTEAPRGAMKILGLFYAALYYPPLAACATAGHTAAHLGSLTSLWA 173

OY 182 HLGVOVMORAECPQVPKI 199
Db 174 HLGVOVMORAECPQVPKI 191

RESULT 9

US-09-901-812-5
Sequence 5, Application US/09901812
Patent No. US20020173461A1
GENERAL INFORMATION:
APPLICANT: Pennica, Diane
APPLICANT: Polakis, Paul
APPLICANT: Szeto, Wayne
TITLE OF INVENTION: UPREGULATION OF TUMOR ANTIGENS TO
FILE REFERENCE: GENENT, 083A
CURRENT APPLICATION NUMBER: US/09/901,812
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/228,914
PRIOR FILING DATE: 2000-08-29
PRIOR APPLICATION NUMBER: 09/759,056
PRIOR FILING DATE: 2001-01-11
PRIOR APPLICATION NUMBER: 60/175,849
PRIOR FILING DATE: 2000-01-13
PRIOR APPLICATION NUMBER: 60/197,089
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 658
TYPE: PRT
ORGANISM: Homo sapiens
US-09-901-812-5

Query Match 93.0%; Score 986.5; DB 9; Length 658;
Best Local Similarity 94.4%; Pred. No. 6.1e-88;
Matches 187; Conservative 1; Mismatches 1; Indels 9; Gaps 1;

OY 2 SOPAGNQTSPGATEDYSGSWYIDEPDQGXELQPEGEVPSCHTSTIPGLYHACLASISL 61
Db 3 SOPAGNQTSPGATEDYSGSWYIDEPDQGXELQPEGEVPSCHTSTIPGLYHACLASISL 62
OY 62 VLLIAMLVRRRLMPDCVGRGPGLPSPVDFLAGDRPRAVPAVFMVLLSLCLLPD 121
Db 63 VLLIAMLVRRRLMPDCVGRGPGLPSPVDFLAGDRPRAVPAVFMVLLSLCLLPD 113
OY 122 ALPFLTLASAPSDGKTEAPRGAMKILGLFHYAALYYPPLAACATAGHTAHLIGSTLSWA 181
Db 114 ALPFLTLASAPSDGKTEAPRGAMKILGLFHYAALYYPPLAACATAGHTAHLIGSTLSWA 173
OY 182 HLGVOVMORAECPQVPKI 199
Db 174 HLGVOVMORAECPQVPKI 191

RESULT 10

US-09-816-653A-7
Sequence 7, Application US/09816653A
Patent No. US20030021788A1
GENERAL INFORMATION:
APPLICANT: Pennica, Diane
APPLICANT: Rastelli, Luca
TITLE OF INVENTION: NOVEL HUMAN STRA6-LIKE PROTEIN AND NUCLEIC ACIDS ENCODING THE SAME
FILE REFERENCE: 10716/57
CURRENT APPLICATION NUMBER: US/09/816,653A
CURRENT FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: 60/191,532
PRIOR FILING DATE: 2000-03-23
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 670

TYPE: PRT
ORGANISM: Mus musculus retinoic acid-responsive protein (Stra6) mRNA, complete cds
US-09-816-653A-7

Query Match 65.7%; Score 697; DB 9; Length 670;
Best Local Similarity 68.6%; Pred. No. 8.5e-60;
Matches 133; Conservative 18; Mismatches 41; Indels 2; Gaps 1;

OY 6 GNQTSFGATEDYSGSWYIDEPDQGXELQPEGEVPSCHTSTIPGLYHACLASISL 65
Db 9 GSQTSFGATEDYSGSWYIDEPDQGXELQPEGEVPSCHTSTIPGLYHACLASISL 66
OY 66 LAMLVRRRLMPDCVGRGPGLPSPVDFLAGDRPRAVPAVFMVLLSLCLLPD 125
Db 67 LAMLVRRRLMPDCVGRGPGLPSPVDFLAGDRPRAVPAVFMVLLSLCLLPD 126
OY 126 LTLASAPSDGKTEAPRGAMKILGLFHYAALYYPPLAACATAGHTAHLIGSTLSWA 185
Db 127 LTLASAPSDGKTEAPRGAMKILGLFHYAALYYPPLAACATAGHTAHLIGSTLSWA 186
OY 186 QVMORAECPQVPKI 199
Db 187 QVMORAECPQVPKI 200

RESULT 11

US-09-925-297-738
Sequence 738, Application US/09925297
Patent No. US20020081659A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA105
CURRENT APPLICATION NUMBER: US/09/925,297
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05989
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 928
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 738
LENGTH: 624
TYPE: PRT
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (188)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (192)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-738

Query Match 7.9%; Score 84; DB 10; Length 624;
Best Local Similarity 22.7%; Pred. No. 3;
Matches 39; Conservative 20; Mismatches 73; Indels 40; Gaps 7;

OY 7 NOTSPGATEDYSGSWYIDEPDQGXELQPEGEVPSCHTSTIPGLYHACLASISL 61
Db 112 NNSPALGTFPPAPMPPGPPTNFSLELEPRGQCVARAEGPSPTAILIGCTVAIIL 171
OY 62 VLLIAMLVRRRLMPDCVGRGPGLPSPVDFLAGDRPRAVPAVFMVLLSLCLLPD 93
Db 172 LLLIILALMLRRLMRRLSKERRVLEELTVHLSVPGDTLLINRRPGPREPPVQERRP 231
OY 94 AGDRPRAVPAVFMVLLSLCLLPDDELPLFLTLASAPSDGKTEAPRGAM 145
Db 232 RGNPHSAPC-----VPNGSALLSNPAVRLLATYARPPRG-----PPTPAW 276

```
RESULT 12
US-09-223-490-4
; Sequence 4, Application US/09223490
; Patent No. US20020147325A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie R.
; APPLICANT: Scadden, David T.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Baron, Will F.
; TITLE OF INVENTION: Protein Tyrosine Kinases
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/223,490
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/170,558
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 854C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-09-223-490-4

Query Match
Best Local Similarity 22.7%; Pred. No. 5.9;
Matches 39; Conservative 20; Mismatches 73; Indels 40; Gaps 7;

QY 7 NOTSPGATEDYSGWYIDEPQ---GGXELQPEGEVPSCH-TSIPGLYHACLASLSTL 61
DB 370 NNSPALGTFPPAPWPGPPTNFSLEBRGQGPVAKPEGSTALICLVAIILL 429
QY 62 VILLMLV-----RRQLMPD-----CVRGPGGLSPVDL---- 93
DB 430 LLLIALLMLRMLRRLSKARRVLEELTVHLVSPGDTILINRPGPREPPYQEPRP 489
QY 94 AGDRPRAVAAMVFMVLLSLCLLPEDDLPLTLASAPSQGKTEAPRGAW 145
DB 490 KSNPPLSAPC---VNGSALLLSNPAYRLLLATYARPRPGG---PPTPAW 534

RESULT 13
US-09-808-880-4
; Sequence 4, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Kriehnkant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
```

```
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808, 880
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3519
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; US-09-808-880-4

Query Match
Best Local Similarity 23.5%; Pred. No. 82;
Matches 54; Conservative 25; Mismatches 70; Indels 81; Gaps 12;

QY 11 PGATEDYSGWYIDEPGCGXELQPEGEVPSCHTSIPGLYHACLASLSTLVLLMLV 70
DB 2642 PAALDEVVGGGLFVYVRG-----LADGVARVVAATA 2676
QY 71 R-----RRLMPDCVGRGPLSPVDPLAGDRPR-----AVPA--AVFMVL 109
DB 2677 RCGEVSVELDPTPRDRAVAEVAAGR-GVSGVVSFLSMDRRRSEHPVPAGLAASVL 2735
QY 110 LSSLC-----LLLPDDALPFLTLASAPSQGKTEAPRCA--W---KILGLFH-- 152
DB 2736 AALVDLGRVGEGRPLMLVTBDA-----VWAGPSDAGAVIDPVQAQWGFGRVLGLEHPE 2790
QY 153 --YVALVYPLAA---CATAGHTAHLGSLTSMVHL-----GVQVWR 190
DB 2791 LMGGLIDLPVAPBPGSTCDHTYADLTLATVVASGFEQOVAVRGSGVWR 2840

RESULT 14
US-09-738-626-6358
; Sequence 6358, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 6358
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
```

US-09-738-626-6358

Query Match 7.0%; Score 74.5; DB 9; Length 208;
Best Local Similarity 26.5%; Pred. No. 6.7; 71; Indels 43; Gaps 11;
Matches 50; Conservative 25; Mismatches 79; Indels 43; Gaps 11;

QY 21 SWYID-----EPGGXELQPEG---EVPSCHTSI---PPGLYHACLASLILVL 63
DB 26 AMFVDTSLADLINTSRPACVELLSQKASLNLISLCSQSLPQPPRLTSAATTSMSPRTL 85
QY 64 L--LLAMLVRRRLQIWPCCVGRPGCLPSPVDFLAGDRRAVPAVFMVILSLCLLPDED 121
DB 86 LKNVSAKLIVRSCAMPPLLFTRILRSPVT--VASSLEPMLSSASMKWAPS-CLVMSKSLP 143
QY 122 ALPPLTLASAPSDGKTEAPRGAKIL--GLFHYALVYPLAACATAGHTAHLIGSTL- 178
DB 144 QIPPAT-----GLMKAKODLCNQALT---SNSCAT--GSPALNLAGTRIP 184
QY 179 SWAHLGVQV 187
DB 185 AMSHQLCQV 193

RESULT 15

US-09-925-301-1349
; Sequence 1349, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1349
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (137)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (142)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1349

Query Match 7.0%; Score 74; DB 10; Length 326;
Best Local Similarity 28.0%; Pred. No. 13;
Matches 58; Conservative 16; Mismatches 79; Indels 54; Gaps 12;

QY 38 EVPSCHTST-----PPGLYHACLASLILVLMLVRRROLMPDCVRPGLP 87
DB 39 DVAACPGSLDCLKRRARCPGA-HACGCLQPFQEDQGLCVPMRRPFGGRPOPRIE 97
QY 88 SPVDFLA-----GDRPR-AVPAAV-FMVLSSLCLLPDEDALPFLT-- 127
DB 98 DEIDFLAQLARKESGHTPPLPKDRQRLPEPATLGFSAKGGGLKGLPSTGPTPTPH 157
QY 128 --LASAPSDG---KTEAPRGAKILGLFHYALVYPLAACATAGHTAHLIGSTLSWPH 182
DB 158 TSLGSPVSSDPYHMSPLEBRG-----GGDGLALVLIILAF-C-VAG--AAALSVASLCWCR 209
QY 183 L--GVQVMQRAE-----CPQVPKI 199
DB 210 LQREIRLTKADATATAKAPGSPAAPRI 236

Search completed: March 13, 2003, 11:37:57
Job time : 12.355 secs

•
•
•
•
•

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 13, 2003, 11:26:32 ; Search time 20.8874 Seconds
(without alignments)
1963.067 Million cell updates/sec

Title: US-09-816-653A-2

Sequence: 1 MSQPPAGNQTSPGATEDPYSYG.....MAHLGVQVMQRAECQVPKRI 199

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL.21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virins:*
16: sp_bacterioplasmid:*
17: sp_archaeoplasmid:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1048	98.8	667	Q9BX79	Q9BX79 homo sapien
2	1048	98.8	667	Q8TB21	Q8TB21 homo sapien
3	986.5	93.0	658	Q9BX78	Q9BX78 homo sapien
4	697	65.7	670	Q70491	Q70491 mus musculus
5	90	8.5	263	Q8SUI0	Q8SUI0 encephalito
6	89.5	8.4	750	Q9YQ39	Q9YQ39 pseudorabies
7	87.5	8.2	336	Q94LD3	Q94LD3 oryza sativ
8	87.5	8.2	413	Q8XRV5	Q8XRV5 ralsionia s
9	86.5	8.2	757	Q9A2Z5	Q9A2Z5 caulobacter
10	85.5	8.1	569	Q8Y1X8	Q8Y1X8 ralsionia s
11	84.5	8.0	382	Q93505	Q93505 bovine hept
12	84.5	8.0	730	Q93OX7	Q93OX7 corynebacte
13	83	7.8	1172	Q9ZBY5	Q9ZBY5 streptomyce
14	82	7.7	208	Q36437	Q36437 laesaea sp.
15	82	7.7	2641	Q9BXD4	Q9BXD4 homo sapien
16	81.5	7.7	9507	Q9EWAI	Q9EWAI streptomyce

17	81	7.6	278	11	Q91YD1	Q91YD1 mus musculus
18	81	7.6	458	2	Q9F2A0	Q9F2A0 chromacium
19	80.5	7.6	120	17	Q8ZUY6	Q8ZUY6 pyrobaculum
20	80.5	7.6	241	2	Q8VPR4	Q8VPR4 micrococcus
21	80.5	7.6	385	16	Q914M7	Q914M7 pseudomonas
22	79.5	7.5	151	11	Q9CYX5	Q9CYX5 mus musculus
23	79.5	7.5	249	2	Q57382	Q57382 escherichia
24	79.5	7.5	664	16	Q9F2V5	Q9F2V5 streptomyce
25	79	7.4	433	12	Q98590	Q98590 prospect hl
26	79	7.4	579	10	Q9FMC8	Q9FMC8 oryza sativ
27	79	7.4	1312	3	Q8WZV2	Q8WZV2 neurospora
28	78.5	7.4	447	4	Q9Y220	Q9Y220 homo sapien
29	78.5	7.4	477	16	Q91660	Q91660 pseudomonas
30	78.5	7.4	601	16	Q8Y3B6	Q8Y3B6 ralsionia s
31	77.5	7.3	156	10	Q8R258	Q8R258 oryza sativ
32	77.5	7.3	439	16	Q9F2M1	Q9F2M1 streptomyce
33	77.5	7.3	476	16	Q8XZ18	Q8XZ18 ralsionia s
34	77.5	7.3	1311	6	Q9BDH6	Q9BDH6 equus caball
35	77	7.3	521	16	Q8YJ03	Q8YJ03 brucella me
36	76.5	7.2	280	2	Q8RTX3	Q8RTX3 xanthomonas
37	76.5	7.2	311	3	Q8WZM5	Q8WZM5 trichoderma
38	76.5	7.2	406	10	Q9FRB9	Q9FRB9 oryza sativ
39	76	7.2	128	4	Q96DK3	Q96DK3 homo sapien
40	76	7.2	203	8	Q33884	Q33884 arabidopsis
41	76	7.2	394	16	Q53924	Q53924 streptomyce
42	76	7.2	539	10	Q8S7M1	Q8S7M1 oryza sativ
43	76	7.2	666	16	Q91143	Q91143 pseudomonas
44	76	7.2	709	16	Q9RYK7	Q9RYK7 deinococcus
45	75.5	7.1	360	11	Q8R4S8	Q8R4S8 rattus norv

ALIGNMENTS

RESULT 1

ID	Q9BX79	PRELIMINARY;	PRT;	667 AA.
AC	Q9BX79; Q9H908;			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 21, Last annotation update)			
DE	SRRA6 isoform 1.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Szeleto W., Jiang W., Tice D.A., Rubinfeld B., Hollingshead P.G.,			
RA	Fong S.E., Dugger D.L., Pham T., Yansura D.G., Wong T.A.,			
RA	Grimaldi J.C., Corpuz R.T., Singh J.S., Frantz G.D., Devaux B.,			
RA	Crowley C.W., Schwall R.H., Eberhard D.A., Rastelli L., Polakis P.,			
RA	Pennica D.;			
RT	"Overexpression of the retinoid acid-responsive gene Stra6 in human			
RT	cancers and its synergistic induction by Wnt-1 and retinoic acid.";			
RL	Cancer Res. 0:0-0(2001).			
RN	[2]			
RP	SEQUENCE OF 108-667 FROM N.A.			
RA	Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,			
RA	Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,			
RA	Takahashi M., Chiba Y., Iehida S., Murakawa K., Ono Y., Takiguchi S.,			
RA	Watanabe S., Kimura K., Murakami K., Iehi S., Kawai Y., Saito K.,			
RA	Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuo Y.,			
RA	Ninomiyama K., Iwayanagi T.;			
RT	"NEO human cDNA sequencing project.";			
RT	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF522728; AAK30289.1; -			
DR	EMBL; AK022603; BAB14122.1; -			
FT	VARIANT 527 527 M -> I			
SQ	SEQUENCE 667 AA; 73502 MW; D20840A46998BA2E CRC64;			

Query Match

98.8%; Score 1048; DB 4; Length 667;

Best Local Similarity 99.0%; Pred. No. 3.3e-95;
Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPGLYHACIASISIL 61
DB 3 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPGLYHACIASISIL 62
OY 62 VLLLLAMLVRRRQLMPDCVGRPGLPSPVDFLAGDRPRAVPAVFMVLLSSLCILLPDED 121
DB 63 VLLLLAMLVRRRQLMPDCVGRPGLPSPVDFLAGDRPRAVPAVFMVLLSSLCILLPDED 122
OY 122 ALPFLTLASAPSDQKTEAPRGAMKILGLFYAALYPLAACATAGHTAAHLGSLTSMWA 181
DB 123 ALPFLTLASAPSDQKTEAPRGAMKILGLFYAALYPLAACATAGHTAAHLGSLTSMWA 182
OY 182 HLGQVQWQRAECPOVPKI 199
DB 183 HLGQVQWQRAECPOVPKI 200

RESULT 2

OY 08TB21 PRELIMINARY; PRT; 667 AA.
ID 08TB21;
AC 08TB21;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE Hypothetical 73.5 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC025256; AAH25256.1; --
KW Hypothetical protein.
SQ SEQUENCE 667 AA; 73532 MW; 120685F4635CB0E5 CRC64;

Query Match 99.8%; Score 1048; DB 4; Length 667;
Best Local Similarity 99.0%; Pred. No. 3.3e-95;
Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPGLYHACIASISIL 61
DB 3 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPGLYHACIASISIL 62
OY 62 VLLLLAMLVRRRQLMPDCVGRPGLPSPVDFLAGDRPRAVPAVFMVLLSSLCILLPDED 121
DB 63 VLLLLAMLVRRRQLMPDCVGRPGLPSPVDFLAGDRPRAVPAVFMVLLSSLCILLPDED 122
OY 122 ALPFLTLASAPSDQKTEAPRGAMKILGLFYAALYPLAACATAGHTAAHLGSLTSMWA 181
DB 123 ALPFLTLASAPSDQKTEAPRGAMKILGLFYAALYPLAACATAGHTAAHLGSLTSMWA 182
OY 182 HLGQVQWQRAECPOVPKI 199
DB 183 HLGQVQWQRAECPOVPKI 200

RESULT 3

OY 09BX78 PRELIMINARY; PRT; 658 AA.
ID 09BX78;
AC 09BX78;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE STRA6 isoform 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Szeto W., Jiang W., Tice D.A., Rubinfield B., Hollingshead P.G.,
RA Fong S.E., Dugger D.L., Pham T., Yansura D.G., Wong T.A.,
RA Grimaldi J.C., Corpuz R.T., Singh U.S., Frantz G.D., Devaux B.,
RA Crowley C.W., Schwall R.H., Eberhard D.A., Raetzell L., Polakis P.,
RA Pennica D.;
RT "Overexpression of the retinoic acid-responsive gene STRA6 in human
cancers and its synergistic induction by Wnt-1 and retinoic acid.";
RT Cancer Res. 0:0-0(2001).
DR EMBL: AF352729; AAC30290.1; --
FT VARIANT 518 518
SQ SEQUENCE 658 AA; 72582 MW; 5405281727F647DD CRC64;

Query Match 93.0%; Score 986.5; DB 4; Length 658;
Best Local Similarity 94.4%; Pred. No. 3.9e-89;
Matches 187; Conservative 1; Mismatches 1; Indels 9; Gaps 1;

OY 2 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPGLYHACIASISIL 61
DB 3 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPGLYHACIASISIL 62
OY 62 VLLLLAMLVRRRQLMPDCVGRPGLPSPVDFLAGDRPRAVPAVFMVLLSSLCILLPDED 121
DB 63 VLLLLAMLVRRRQLMPDCVGRPGLPSPVDFLAGDRPRAVPAVFMVLLSSLCILLPDED 113
OY 122 ALPFLTLASAPSDQKTEAPRGAMKILGLFYAALYPLAACATAGHTAAHLGSLTSMWA 181
DB 114 ALPFLTLASAPSDQKTEAPRGAMKILGLFYAALYPLAACATAGHTAAHLGSLTSMWA 173
OY 182 HLGQVQWQRAECPOVPKI 199
DB 174 HLGQVQWQRAECPOVPKI 191

RESULT 4

ID 070491 PRELIMINARY; PRT; 670 AA.
AC 070491;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE Retinoic acid-responsive protein.
GN STRA6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95377533; PubMed=7649373;
RA Bouillet P., Oulad-Abdelghani M., Vicaire S., Garnier J.M.,
RA Schunhauer B., Dolle P., Chambon P.;
RT "Efficient cloning of cDNAs of retinoic acid-responsive genes in p19
embryonal carcinoma cells and characterization of a novel mouse gene,
RT Stral (mouse LERK-2/Bp182).";
RL Dev. Biol. 170:420-433(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97346723; PubMed=9203140;
RA Bouillet P., Sapin V., Chazaud C., Messadeg N., Decimo D., Dolle P.,
RA Chambon P.;
RT "Developmental expression pattern of Stra6, a retinoic acid-responsive
RT gene encoding a new type of membrane protein.";
RL Mech. Dev. 63:173-186(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA Bouillet P.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF062476; AAC16016.1; --
MGD: MGI:107742; Stra6.
SQ SEQUENCE 670 AA; 73774 MW; 6AC6E9F3BE16A4F7 CRC64;

Query Match 65.7%; Score 697; DB 11; Length 670;
 Best Local Similarity 68.6%; Pred. No. 1,6e-60;
 Matches 133; Conservative 18; Mismatches 41; Indels 2; Gaps 1;

QY 6 GNOTSPGATEDYSYSGWIDEPDQGXELQPEGEVPSCHTSIPGLYHNACLASISLVLL 65
 DB 9 GSGTSSGVTDDYS--SWYIEPFLGAEVQPEGVPLCOLTAPALHACLASISFLVLL 66
 QY 66 LAMVRRRLMPDCVRGPGLPSPVDFLAGRPRAPVAVMVLSSICLLPDEDLRF 125
 DB 67 LALLVRRRLMPDCVRGPGLPSPVDFLAGRPRAPVAVMVLSSICLLPDEDLRF 126
 QY 126 LTLASAPQDQKTEAPRGAMKILGFHYAALYPLAACATAGHTAAHLSTLSMAHLGV 185
 DB 127 LNLTAASSPDDEMETSRKPMKLLALYPLALYPLAACASGHQAFLGLTSLMAHGV 186
 QY 186 QVMQRAECPOVPKI 199
 DB 187 QVMQRAECPOVPKI 200

RESULT 5
 Q8SU10 PRELIMINARY; PRT; 263 AA.
 AC 08SU10;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein ECU08_2060.
 GN ECU08_2060.
 OS Encephalitozoon cuniculi.
 OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
 NCBI_TaxID=6035;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RA Katinka M.D., Duprat S., Cornille E., Metenier G., Thomarat F., Premsier G., Barbe V., Peyretailade E., Brottier P., Wincker P., Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M., Weissenbach J., Vivares C.P.;
 RA "Genome sequence and gene compaction of the eukaryote parasite RT Encephalitozoon cuniculi."
 RT Nature 414:450-453(2001).
 DR EMBL; AL590448; CAD26508.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 263 AA; 30340 MW; 3254A0CC3469A24B CRC64;

Query Match 8.5%; Score 90; DB 5; Length 263;
 Best Local Similarity 23.3%; Pred. No. 0.56;
 Matches 34; Conservative 30; Mismatches 52; Indels 30; Gaps 6;

QY 50 LYHACLASISIL-VLLMLVRRRLMPDCVRGPGLPSPVDFLAGRPRAPVAA-- 104
 DB 99 LNLTLTFTPSILSILSIILINRRNDLFFPSVILBPMPLTLLSTSCRLVPGQIGF 158
 QY 105 -----VFWLSSSLCLL-----LPDEDLRFPLTASAPS-----ODGKTEAPR 142
 DB 159 IDPGINIFIDILSLCLVSLTLCIEKQDLCITAISSALTIVRLKKYLSKSPPT 218
 QY 143 GAWKI--LGLFHYAALYPLAACAT 165
 DB 219 APWRVAIFVLIFSFWFIYVLAAGS 244

RESULT 6
 Q9YQ39

ID Q9YQ39 PRELIMINARY; PRT; 750 AA.
 AC Q9YQ39;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE UL47 putative tegument protein.
 GN UL47.
 OS Pseudorabies virus (strain Kaplan) (PRV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirinae.
 NCBI_TaxID=33703;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KAPLAN;
 RX MEDLINE=99239778; PubMed=10225272;
 RA Bras F., Dezelee S., Simonet B., Nguyen X., Vende P., Flaman A., Masse M.J.;
 RA "The left border of the genomic inversion of pseudorabies virus RT contains genes homologous to the UL46 and UL47 genes of Herpes Simplex RT virus type 1, but no UL45 gene."
 RL Virus Res. 60:29-40(1999).
 DR EMBL; AJ010303; CA009073.1; -
 DR InterPro; IPR005029; Herpes UL47.
 DR Pfam; PF03362; Herpes UL47; 1.
 SQ SEQUENCE 750 AA; 80379 MW; FA82C9B18FBD8F22 CRC64;

Query Match 8.4%; Score 89.5; DB 12; Length 750;
 Best Local Similarity 26.1%; Pred. No. 2;
 Matches 52; Conservative 22; Mismatches 58; Indels 67; Gaps 13;

QY 6 GNOTSPGATE--DYSYSGWIDEPDQGXELQPEGEVPSCHTSIPGLYHNACLASISLVLL 63
 DB 244 GDDSPAPLERPDAAVAYSWRMDPDG-----TFL 272
 QY 64 LLLAMVRRRLMPDCVRGPGLP-SPVDFLAGRPRAPVAVMVLSSICLLPDEDA 122
 DB 273 AMPSPVALRR-----RGPRGRVGPDPILR--RAPALRTTPQAASALVSRDODA 323
 QY 123 ----LPFLTL-----ASAPSDG--KTEAP--RGAWK-----LGLFHYAALYPLAA 162
 DB 324 WEALVPLHLQMHSGWGSAPPSRGVAVATPAPETRGVWRALROALH-HRAMFLAPLOS 382
 QY 163 CATAGHTAAHLGSLTSSWA 181
 DB 383 -STVG--APALAGALAPA 398

RESULT 7
 Q94LD3 PRELIMINARY; PRT; 336 AA.
 AC Q94LD3;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical 36.5 kDa protein.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPONBARE;
 RA Buell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Ganeberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M., Tsilix T., Riggs F., Hsiao J., Ziemann V., Blunt S., Pai G., VanAken S.E., Uterback T.R., Feldjym T.V., Quackenbush J., Salzberg S.L., White O., Frazer C.M.;
 RA "Oryza sativa chromosome 3 BAC OSJNBa0026A15 genomic sequence."
 RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC084404; AAK50595.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 336 AA; 36535 MW; DBA6FEFB509583C9 CRC64;

Query Match 8.2%; Score 87.5; DB 10; Length 336;
 Best Local Similarity 34.7%; Pred. No. 1.3; Mismatches 42; Indels 11; Gaps 5;
 Matches 35; Conservative 13; Mismatches 42; Indels 11; Gaps 5;

QY 43 HTSIPGLYACIASILVLLMLV---RRQLMPDCVGRPGL-PSVPDLFAGDR 97
 DB 211 HRRIRSGVFHVAIVSVQLLAVPRLLAGRRRRRRRRPVPVGRGVRSVAFSSVVR 270

QY 98 PRAV---PAAVFMVLLSSCLLPDDEL-PLUTLASAPS 133
 DB 271 VROYCRSPVVF-VLASASSLVPAASRLPRIAAEVVS 310

RESULT 8
 OXRV5 PRELIMINARY; PRT; 413 AA.

AC OXRV5; PRELIMINARY; PRT; 413 AA.

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DE Putative MFS transporter transmembrane protein.

GN RSP0726 OR RS01724.

OS *Ralstonia solanacearum* (Pseudomonas solanacearum).

OC *Ralstonia solanacearum* (Pseudomonas solanacearum).

OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;

OC Ralstonia.

OX NCBI_TaxID=305;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=GM11000;

RX MEDLINE=21681879; PubMed=11823852;

RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenc S., Arlat M., Billault A., Brotier P., Camus J.C., Cartolico L., Chandel M., Choise N., Claudel-Renard C., Cunac S., Demange N., Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T., Sigler P., Thebaud P., Whalen M., Wincker P., Levy M., Weissenbach J., Boucher C.A.;

RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*."

RL Nature 415:497-502(2002).

DR EMBL: AL646080; CAD17877.1;

KW Plasmid: Complete proteome.

SO SEQUENCE 413 AA; 43486 MW; F4C0B6453C037CCC CRC64;

Query Match 8.2%; Score 87.5; DB 16; Length 413;
 Best Local Similarity 24.1%; Pred. No. 1.6; Mismatches 75; Indels 45; Gaps 8;
 Matches 47; Conservative 28; Mismatches 75; Indels 45; Gaps 8;

QY 1 MSOPAGNOTSPGATEDSYSGWYIDEPQGXLOPEGEVPSCHTSIPGLYHA---CLAS 57
 DB 134 LTESGREGMAAGVNRALTYASTLIPVGGSL-----IFIDAVALLCCSAV 181

QY 58 LSLVLLMLVRRRLMPDCVGRPGL-PSVPDLFAGDR-----PRAVPA---A 104
 DB 182 LALSAILLRSSIFR-----CQSSRQNPAGKMDPIGIRKTLSTYGGSPKALDVLILA 234

QY 105 VPMVLLSSCLLP-DEDALPPLTLASAPSODGKTEAPRGAMKILGFHYALVYPLAAC 163
 DB 235 LFNFLAPLTLVIRPMHVKSLPFTSLSTL-----GLSEASLIGAGVCG-----ALPFSRLRT 285

QY 164 ATAGHTAHLGSLT 178
 DB 286 GSPKHSILVACASL 300

RESULT 9
 O9AAZ5 PRELIMINARY; PRT; 757 AA.

AC O9AAZ5; PRELIMINARY; PRT; 757 AA.

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Beta-N-acetylhexosaminidase, putative.

GN CC0447;
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 19089 / CB15;

RX MEDLINE=21173698; PubMed=11259647;

RA Nierman W.C., Feldberg T.V., Laub M.T., Paulsen I.T., Nelson K.E., Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Potocka I., Nelson W.C., Newton A.S., Stephens C., Phadke N.D., Ely B., Deboy R.T., Dodson R.J., Durkin A.S., Gehm M.B., Hart D.H., Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K., Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O., Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

RT "Complete genome sequence of *Caulobacter crescentus*."

RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

DR EMBL: AE005718; AAK22434.1;

DR HSSP: P06865; IQBC.

DR TIGR: CC0447; .

DR InterPro: IPR01540; GH_20.

DR Pfam: PF00728; Glyco_hydro_20; 1.

DR Pfam: PF02838; Glyco_hydro_20b; 1.

DR PRINTS: PR00738; GLHYDRASE20.

KW Complete proteome.

SO SEQUENCE 757 AA; 82171 MW; 7E24C4E7B173803 CRC64;

Query Match 8.2%; Score 86.5; DB 16; Length 757;
 Best Local Similarity 30.6%; Pred. No. 4; Mismatches 49; Conservative 16; Mismatches 60; Indels 35; Gaps 9;

QY 62 VLLIAMVRR-RQLMPDCVGRPGL-PSVPDLFAGDR-----RAVPA 103
 DB 65 VALQADIDIOSRGRPPVVEGPPADAIY--LTREBPAGEAYKLDINAKATIAAKRA 122

QY 104 AVFMVLLSSCLLPDDEL-PLUTLASAPSODGKTEAPRGAMKILGFHYALVY---P 159
 DB 123 GLFGAMSLMGLATPDEAKGVALLAISIE-----DAPPRAMR--GLMVSARHYSLDT 175

QY 160 LAACATGHTAHLGSLTSLVAILGVQVWQRAECPPQPKI 199
 DB 176 LKAVIDA--MAAHKL-NTFHWHLVDQGM-RLEIKKPKL 211

RESULT 10
 O8Y1X8 PRELIMINARY; PRT; 569 AA.

AC O8Y1X8; PRELIMINARY; PRT; 569 AA.

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DE Hypothetical transmembrane protein RSC0559.

GN RSC0559 OR RS04906.

OS *Ralstonia solanacearum* (Pseudomonas solanacearum).

OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;

OC Ralstonia.

OX NCBI_TaxID=305;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=GM11000;

RX MEDLINE=21681879; PubMed=11823852;

RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenc S., Arlat M., Billault A., Brotier P., Camus J.C., Cartolico L., Chandel M., Choise N., Claudel-Renard C., Cunac S., Demange N., Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T., Weissenbach J., Thebaud P., Whalen M., Wincker P., Levy M.,

RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*."

RL Nature 415:497-502(2002).

DR EMBL: AL646060; CAD14089.1;

KW Hypothetical protein; Complete proteome.

SO SEQUENCE 569 AA; 62435 MW; C1AB4805E1A843EB CRC64;

Query Match 8.1%; Score 85.5; DB 16; Length 569;
 Best Local Similarity 24.1%; Pred. No. 3.6; Indels 60; Gaps 9;
 Matches 49; Conservative 21; Mismatches 60; Indels 73; Gaps 9;

49 GLYHACLASISILVLLMLVRRRLMPDCVRGRPGLP-----SPVDFLAG----- 95
 56 GLPKVYLAFLALIGLVIVLVV-----ATPLNPFPTFAIVFLGTVAVCGIGA 104
 96 ---DRPRAVPAVFMVLISSL-----CLLPDEDALPFLTLASAPSDGKTEPRGAWK 146
 105 RCRDVPGLVLESIAVAIVIGLITVAIECLHLFRVSGLPINMISIMP-----TGAGRMMWG 159
 147 ILGLFHYALLY--YFLAACATAGHT-----AAHLIG-----STLSMAHL----- 183
 160 NINOPNHVATYLAFLGACPLGSTRRRRYMAPLAIALALLGHALTYVSRMSHLVIVG 219
 184 -----GVQWQRAECP 194
 220 GVAGLAMSAERGAARWRACVP 242

RESULT 11

039505 PRELIMINARY; PRT; 382 AA.

AC 039505; (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Glycoprotein I.
 GN US7.
 OS Bovine herpesvirus 1.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirinae.
 OX NCBI_TaxID=10320;
 RN RP SEQUENCE FROM N.A.
 RC STRAIN=JURA;
 RA Goltz M., Buhk H.J., Broll H., Lewin M., Mankertz A., Boerner B.,
 RA Borchers K., Weigelt W.
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE OF 372-382 FROM N.A.
 RC STRAIN=JURA;
 RA Schwytzer M.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=JURA;
 RA Schwytzer M., Paces V., Letchworth G.J., Misra V., Buhk H.J.,
 RA Lowery D.E., Simard C., Bello L.J., Thiry E., Vizek C.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=COOPER;
 RA Goltz M., Buhk H.J., Schwytzer M.;
 RT "Nucleotide sequence of the HindIII O and K fragments located in the
 RT US region of the bovine herpesvirus 1 genome";
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ004801; CA06146.1; -
 DR EMBL; Z98199; CAB10890.1; -
 DR InterPro: IPR002874; Herpes_g1.
 DR Pfam: PF01688; Herpes_g1.
 SQ SEQUENCE 382 AA; 39596 MW; 0F06C95CD2581EC2 CRC64;

Query Match

Best Local Similarity 8.0%; Score 84.5; DB 12; Length 382;
 Matches 37; Conservative 12; Mismatches 38; Indels 47; Gaps 6;

4 PAKNGTSPGATEDSYSGSWITDEPGGKXELQPEGEVSCHTSIPLGILYHACLAS-----LS 59
 258 PAKITAPASPSRAFA-----AAPAAAQPADTRPA-----RFRQLASILVPLC 304

QY 60 ILVLLIAM-----LVRRRL-----WPDVCRGRPLPSPV 90
 DB 305 VLVLLIALCAATVNCALRRLLPCSRVYKPRTCATCGGTCAGRPPCGAAPAPATV 364
 QY 91 DPLAGDRPAPAA 104
 DB 365 VAL-GSRPAPPLA 377

RESULT 12

0930X7 PRELIMINARY; PRT; 730 AA.

AC 0930X7; (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 77.3 kDa protein.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
 OC Corynebacterium.
 OX NCBI_TaxID=1718;
 RN RP SEQUENCE FROM N.A.
 RC STRAIN=M0233;
 RA Krubasik P., Kobayashi M., Sandmann G.;
 RT "Expression and functional analysis of a gene cluster involved in the
 RT synthesis of decaprenoxanthin reveals the mechanisms for C50
 RT carotenoid formation."
 RL Eur. J. Biochem. 286:3702-3708(2001).
 DR EMBL; AF159510; AK64297.1; -
 DR InterPro: IPR001092; HLH_basic.
 DR InterPro: IPR000731; HMGCR_patch_STM.
 DR InterPro: IPR000847; HTH_LysR.
 DR InterPro: IPR004859; MMP.
 DR Pfam: PF03176; MMP.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
 DR PROSITE: PS00156; SSD; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 730 AA; 77290 MW; 18A9CAE08418633 CRC64;

Query Match 8.0%; Score 84.5; DB 2; Length 730;
 Best Local Similarity 29.4%; Pred. No. 6;
 Matches 30; Conservative 18; Mismatches 41; Indels 13; Gaps 3;

QY 49 GLYHACLASISILVLL-----LMLVRRRLMPDCVRGRPGLP-----SPVDFLAGDRP 98
 319 GTFPMASALTLPALFLFVGSRVAFWPKRPKPEPVYHGENGIFASGIMSKVAVLVEQHP 378
 DB 99 RAVPAVFMVLISSICT-----LLPDEDALPFLTLASAPSDGK 137
 379 RAIWSTLIVLLGAAPVTLKADGVSGSDVLGSSSEARDQ 420

RESULT 13

092BY5 PRELIMINARY; PRT; 1172 AA.

AC 092BY5; (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative ABC transporter.
 DE CYDOD OR SCO3947 OR SCD78.14.
 GN Streptomyces coelicolor.
 OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D.C., Harris D.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA MEDLINE=97000351; PubMed=843436;
RA Redenbach M., Kleiser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome";
RN [4]
RL Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -1 SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AL014355; CA022219.1; -.
DR HSSP; P13569; INBD.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001140; ABCtransportTM.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR000524; HTH_GNTR.
DR Pfam; PF00664; ABC_membrane_2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR PROSITE; PS00043; HTH_GNTR_FAMILY; UNKNOWN_1.
DR ATP-binding; Transport.
DR KW
SQ SEQUENCE 1172 AA; 122766 MW; 5710CD932ED704D CRC64;

Query Match
Best Local Similarity 7.8%; Score 83; DB 16; Length 1172;
Matches 45; Conservative 12; Mismatches 68; Indels 18; Gaps 5;

QY 55 LASTSLVLLMLVRRQLWPCVGRPGLPSPVDFLAGDRPRAVPAVFMVLLSSLC 114
DB 674 LAAFEAVLGLPLAVOYROR-VRSASERYVEVLDAEPVREPEVPRQAPTPPPVVEGLA 932
QY 115 LLLP--DEDALEPFLTL-----ASAPSDQKTEAPRGAMKILGFHYALVYPLA-- 161
DB 933 ARNPGQDRDALAGLDLTLEQGRVAVVGSPSGKTTL--AQVLLRFLDSAAAGSTTLGCV 989
QY 162 -ACATAGTAHAHLGSLSTSMANL 183
DB 990 DAYALAGDQVRLVGLCAQADAL 1012

RESULT 14
Q36437
ID Q36437 PRELIMINARY; PRT; 208 AA.
AC Q36437
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Cytochrome oxidase III protein.
OS Lasea sp.
OG Mitochondrion.
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Heteroconcha; Veneroida;
OC Galeommatoidea; Laseidae; Lasea.

```

```

OX NCBI_TaxID=32592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KERG;
RA O'Foghil D., Smith M.J.;
RT "Evolution of asexuality in the cosmopolitan marine clam Lasea.";
RN [1]
RL Evolution 49:140-150(1995).
DR EMBL; X78301; CA55106.1; -.
DR InterPro; IPR00298; CytoC_oxdae_III.
DR Pfam; PF00510; COX3; 1.
DR PRODOM; PD000382; CytoC_oxdae_III; 2.
DR PROSITE; PS50253; COX3; 1.
DR Mitochondrion.
SQ SEQUENCE 208 AA; 22785 MW; 05517676CB39576 CRC64;

Query Match
Best Local Similarity 7.7%; Score 82; DB 8; Length 208;
Matches 33; Conservative 24; Mismatches 58; Indels 22; Gaps 5;

QY 60 ILVLLMLVRRQL---WPCVGRGRLSPVDFLAGDRPRAVPAVFMVLLSSLC 116
DB 12 VLKFLIFILGLTMTAMWGDVKEGTYLCHTSLVA---RGLRGMALFIISEVFF 67
QY 117 LPDEPALPFLTLASAPSDQKTEAPRGAMKILGFHYALVYPLAACATAGTAHL-IG 175
DB 68 VSFMAFFHLISGL-----AEQGVPMGIFPINAQVPL-----NTAVLLSSG 113
QY 176 STLSAHLGVGVQWQRAE 192
DB 114 VSVTMAHYAIRWSRAQ 130

RESULT 15
Q9BXD4
ID Q9BXD4 PRELIMINARY; PRT; 2641 AA.
AC Q9BXD4
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE PDZ domain-containing protein AIPC.
GN AIPC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=2182735; PubMed=11289102;
RA Chailb H., Rubin M.A., Mucci N.R., Li L., Taylor J.M.G., Day M.L.,
RA Rhim J.S., Macoska J.A.;
RT "Activated in prostate cancer: A PDZ domain-containing protein highly
RT expressed in human primary prostate tumors.";
RN [1]
RL Cancer Res. 61:2390-2394(2001).
DR EMBL; AF338650; AAK07661.1; -.
DR HSSP; Q14005; I116.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 4.
DR SMART; SM00228; PDZ; 6.
DR PROSITE; PS50106; PDZ; 6.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
SQ SEQUENCE 2641 AA; 280205 MW; CEBFE84582975754 CRC64;

Query Match
Best Local Similarity 7.7%; Score 82; DB 4; Length 2641;
Matches 50; Conservative 19; Mismatches 64; Indels 70; Gaps 9;

QY 11 PGLTEQVSY-GSYVIDEPQGXELQPEGEVPSCHTSIP-----PGLYH- 52
DB 679 PGLPSDFMVGSEDEHDPGSGCSTSEGLPETSSEPGKPRANSLVTLGSHRASGLPFK 738
QY 53 ----ACIASISILVLLMLVRRQLM-----PCCVAGRGRLSPVDF 92

```

```

Db 739 QVTVARQASLPGSPQALRNPLLROKVKCYDANDASDEEFREGDCISLPGALPGPIRP 798
Qy 93 IAGDRPR-----AVPAVFMVILSSLCILLPDEDALP--FTL 128
Db 799 LSEDDPRRVGISSSKGMVDHNOEERPRKTIYSKAIAPILGSSVDL---ESIPEGWDA 855
Qy 129 AS-APSODGKTEAPRGA---WK 146
Db 856 ASYAAANLTDSEAAPKGSFGSNWK 878

```

Search completed: March 13, 2003, 11:29:56
 Job time : 25.8874 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:26:57 ; Search time 6.45611 Seconds
(without alignments)
1278.447 Million cell updates/sec

Title: US-09-816-653A-2
Perfect score: 1061
Sequence: 1 MSQPMAGNQTSPGATEDYSYG.....MAHLGVQVWQRAECQVPK1 199

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85.5	8.1	330	1	Q08102 bovine hep
2	84.5	8.0	335	1	Q924W9 streptomyce
3	84	7.9	913	1	Q08345 h epithelia
4	83	7.8	281	1	P56261 rattus norv
5	80	7.5	433	1	P22047 prospect hi
6	80	7.5	925	1	Q13469 homo sapien
7	79	7.4	910	1	Q63474 rattus norv
8	78.5	7.4	532	1	Q14451 homo sapien
9	78.5	7.4	577	1	Q04648 oenothera b
10	78.5	7.4	3519	1	Q07017 streptomyce
11	78	7.4	281	1	Q15744 homo sapien
12	78	7.4	419	1	Q04903 pisanu sattiv
13	78	7.4	911	1	Q01146 mus musculu
14	77.5	7.3	599	1	Q01647 daucus caro
15	76.5	7.2	874	1	P97481 mus musculu
16	75.5	7.1	655	1	Q01634 canis famli
17	74	7.0	867	1	P51517 samian retr
18	73	6.9	786	1	O50724 mycobacteri
19	72.5	6.8	636	1	P15073 milk cell f
20	72	6.8	1385	1	Q14518 chlamydomon
21	71.5	6.7	1844	1	P28477 turnip yell
22	71	6.7	193	1	P47206 pseudomonas
23	71	6.7	325	1	O9nqx5 homo sapien
24	70.5	6.6	430	1	Q63156 rattus norv
25	70.5	6.6	628	1	P51179 rattus norv
26	70.5	6.6	888	1	P38036 escherichia
27	70	6.6	931	1	Q9wcm3 mus musculu
28	70	6.6	1032	1	P72637 synchocyst
29	70	6.6	1189	1	Q62130 mus musculu
30	69.5	6.6	262	1	O47425 branchiosto
31	69.5	6.6	329	1	P41031 gallus gall
32	69.5	6.6	425	1	O43613 homo sapien
33	69.5	6.6	465	1	Q50186 mycobacteri

34	69.5	6.6	867	1	POL_MPMV	P07572 simian maso
35	69.5	6.6	1204	1	NOS3_BOVIN	P29473 bos taurus
36	69	6.5	3412	1	POLG_TBRSVS	P07720 t genome po
37	68.5	6.5	194	1	FGF4_CHICK	P48804 gallus gall
38	68.5	6.5	273	1	DLX3_NORVI	P53770 notophthalm
39	68.5	6.5	274	1	DLX3_PLEMA	O91284 pleurodeles
40	68.5	6.5	452	1	CPT7_ORYIA	P70085 oryzias lat
41	68.5	6.5	517	1	NIFU_ECOLI	P52647 escherichia
42	68.5	6.5	1174	1	NME3_MOUSE	O01098 mus musculu
43	68.5	6.5	1239	1	MURI_RALSO	Q8xy07 ralsconia s
44	68	6.4	277	1	Y895_MYCTU	Q10554 mycobacteri
45	68	6.4	505	1		

ALIGNMENTS

RESULT 1
VGLI_HSVES STANDARD: PRT: 380 AA.

AC Q08102;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Glycoprotein 1.

GN GI.
OS Bovine herpesvirus type 1.2 (strain ST).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_Taxid=45407;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94167875; Pubmed=8122370;
RA Leung-Tack P., Audonnet J.F., Riviere M.;
RT "The complete DNA sequence and the genetic organization of the short
RT unique region (US) of the bovine herpesvirus type 1 (ST strain).";
RL Virology 199;409-421(1994).
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEINS 1, TO VZV GP1V,
CC AND TO PRV GP63.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: Z23068; CA89605.1; -
DR InterPro: IPR002874; Herpes_g1.
DR Pfam: PF01688; Herpes_g1; 1.
KW Glycoprotein.
FT CARBOHYD 67
SQ SEQUENCE 380 AA; 39910 MW; AEBRIF9BA30D2BD CRC64; (POTENTIAL).

Query Match 8.1%; Score 85.5; DB 1; Length 380;
Best local similarity 27.6%; Pred. No. 1.2; 38; Indels 47; Gaps 6;
Matches 37; Conservative 12; Mismatches

QY 4 PAGNOTSGATEDYSYGSWYIDPQGXLOPEGEVPSCHTSIPPGIYHCLAS---LS 59
DB 256 PAGPTASPAASPSRFSF---AAPAAAAQPADDTA-----RFRQLASILVPLC 302
QY 60 ILVLLLLM-----LVRRRL-----WPDYCRGRGLSPV 90
DB 303 VLVLILLALCAATVNCALRRRLPCSRRVYKPTCAACSGSGTCAGRPCCGAPASAPATV 362
QY 91 DFLAGDRRAVPA 104
DB 363 VAL-GSRKAPPLA 375

RESULT 2

```

TRD2 STRCO
ID ID TRD2 STRCO STANDARD; PRT; 335 AA.
AC 092479;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Anthranilate phosphoribosyltransferase 2 (EC 2.4.2.18).
GN TRPD OR SCO3212 OR SCE8.05C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteriales;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
NCBI_TaxId=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M45;
RX MEDLINE=2196410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrall B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -1- CATALYTIC ACTIVITY: Anthranilate + phosphoribosylpyrophosphate =
CC N-5'-phosphoribosyl-anthranilate + diphosphate.
CC -1- PATHWAY: tryptophan biosynthesis; second step.
CC -1- SIMILARITY: BELONGS TO THE ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL: AL035654; CAB38583.1; -
CC InterPro: IPR000312; Glycosyltransferase-3.
CC Pfam: PF00591; Glycosyltransferase_3; 1.
CC Pfam: PF02885; Glycosyltransferase_3N; 1.
CC ProDom: PD001864; Glycosyltransferase_3; 1.
CC DR TIGRFAMs: TIGR01245; trpD; 1.
CC Tryptophan biosynthesis; transferase; Glycosyltransferase;
CC Complete proteome.
CC KMW SEQUENCE 335 AA; 34536 MW; 2706194E40B2F0D CRC64;
SQ

```

```

DE (Tyrosine-protein kinase CAK) (Cell adhesion kinase) (Tyrosine kinase
DE DDR) (Discoidin receptor tyrosine kinase) (TRK E) (Protein-tyrosine
DE kinase RTK 6) (CD167a antigen).
GN DDRL OR EDDR1 OR CAK OR TRKE OR RTK6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal liver;
RX MEDLINE=9413417; PubMed=8302582.
RA Perez J.L., Shen X., Finkernagel S., Sciorra L., Jenkins N.A.,
RA Gilbert D.J., Copeland N.G., Wong T.W.;
RT "Identification and chromosomal mapping of a receptor tyrosine kinase
RT with a putative phospholipid binding sequence in its ectodomain.";
RL Oncogene 9:211-219(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93296201; PubMed=8390675;
RA Johnson J.D., Edman J.C., Rutter W.J.;
RT "A receptor tyrosine kinase found in breast carcinoma cells has an
RT extracellular discoidin I-like domain.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:5677-5681(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=9713158; PubMed=8977099;
RA Sakuma S., Tada M., Sawa H., Sawamura Y., Shinohara Y., Abe H.;
RT "Receptor protein tyrosine kinase DDR is up-regulated by p53
RT protein.";
RL FEBS Lett. 398:165-169(1996).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=96369017; PubMed=8796349;
RA Playford M.P., Butler R.J., Wang X.C., Katso R.M., Cooke I.B.,
RA Ganeva T.S.;
RT "The genomic structure of discoidin receptor tyrosine kinase.";
RL Genome Res. 6:620-627(1996).
RN [5]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Lung;
RX MEDLINE=96204002; PubMed=8622863;
RA Perez J.L., Jing S.O., Wong T.W.;
RT "Identification of two isoforms of the Cak receptor kinase that are
RT coexpressed in breast tumor cell lines.";
RL Oncogene 12:1469-1477(1996).
RN [6]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Brain, and Keratinocytes;
RX MEDLINE=94043265; PubMed=8226977;
RA di Marco E., Cutuli N., Guerra L., Cancedda R., de Luca M.;
RT "Molecular cloning of trke, a novel trk-related putative tyrosine
RT kinase receptor isolated from normal human keratinocytes and widely
RT expressed by normal human tissues.";
RL J. Biol. Chem. 268:24290-24295(1993).
RN [7]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Ovary;
RX MEDLINE=95151638; PubMed=7848919;
RA Laval S., Butler R., Shelling A.N., Hardy A.M., Poulsen R.,
RA Ganeva T.S.;
RT "Isolation and characterization of an epithelial-specific receptor
RT tyrosine kinase from an ovarian cancer cell line.";
RL Cell Growth Differ. 5:1173-1183(1994).
RN [8]
RP SEQUENCE FROM N.A.
RA Shima S., Tamura G., Oka A., Inoko H.;
RT "Homo sapiens 2,329,817bp genomic DNA of 5p21.3 HLA class I region.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Muscle;

```

RA Strausberg R.; to the EMBL/GenBank/DBJ databases.
CC Submitted (JUG-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY BE INVOLVED IN CELL-CELL INTERACTIONS AND
CC RECOGNITION
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1/CAK I (SHOWN HERE) AND 2/CAK
CC II; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY ONLY DIFFER BY THE
CC ABSENCE OF A 37 RESIDUES SEGMENT.
CC -1- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN MOST ADULT TISSUES
CC AND IS HIGHEST IN THE BRAIN AND LUNG. ABUNDANT IN BREAST CARCINOMA
CC CELLS LINES.
CC -1- DOMAIN: THE GLY/PRO-RICH DOMAINS MAY BE REQUIRED FOR AN UNUSUAL
CC GEOMETRY OF INTERACTION WITH LIGAND OR SUBSTRATES.
CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
CC RECEPTOR SUBFAMILY
CC -1- SIMILARITY: CONTRAINS 1 F5/8 TYPE C DOMAIN.
CC -1- DATABASE: NAME=PROW; NOTE=PROW 1;41-46(2000);
CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/1986613159.g.htm".
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See <http://www.isb-sb.ch/announce>
CC or send an email to license@isb-sb.ch).
CC -----
DR EMBL; L11315; AAA02866.1; -
DR EMBL; L20817; AAA18019.1; -
DR EMBL; U48705; AAC50917.1; -
DR EMBL; X98208; CAA66871.1; -
DR EMBL; X99023; CAA66871.1; JOINED.
DR EMBL; X99024; CAA66871.1; JOINED.
DR EMBL; X99025; CAA66871.1; JOINED.
DR EMBL; X99026; CAA66871.1; JOINED.
DR EMBL; X99027; CAA66871.1; JOINED.
DR EMBL; X99028; CAA66871.1; JOINED.
DR EMBL; X99029; CAA66871.1; JOINED.
DR EMBL; X99030; CAA66871.1; JOINED.
DR EMBL; X99031; CAA66871.1; JOINED.
DR EMBL; X99032; CAA66871.1; JOINED.
DR EMBL; X99033; CAA66871.1; JOINED.
DR EMBL; X99034; CAA66871.1; JOINED.
DR EMBL; L57508; AAB05208.1; -
DR EMBL; L74979; CAA52915.1; -
DR EMBL; L25903; CAA82335.1; -
DR EMBL; AP000511; BAB66318.1; -
DR EMBL; BC000876; AAH08716.1; -
DR EMBL; BC013400; AAH13400.1; -
DR HSSP; P11362; IFGK.
DR Genew; HGNC:2730; DDRL.
DR MIM; 600408; -
DR InterPro; IPRO00719; Euk_pkinase.
DR InterPro; IPRO00421; FAS5.C.
DR InterPro; IPRO02011; RTKkinaseII.
DR InterPro; IPRO01245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR Prodom; PD000001; Euk_pkinase; 1.
DR SMART; SM00231; FAS5C; 1.
DR SMART; SM00219; TYRCK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
DR PROSITE; PS01285; FAS5C_1; 1.
DR PROSITE; PS01286; FAS5C_2; 1.
KW Transferrase; Tyrosine-protein kinase; Glycoprotein; Signal;
KW phosphorylation; Transmembrane; Receptor; ATP-binding; Antigen;
KW Alternative splicing.
FT SIGNAL 1 18 POTENTIAL.

FT	CHAIN	19	913		EPI- EPITHELIAL DISCOIDIN DOMAIN RECEPTOR 1.
ET	DOMAIN	19	416		EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	417	443		POTENTIAL.
FT	DOMAIN	444	913		CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	31	185		F5/8 TYPE C (PHOSPHOLIPID-BINDING, POTENTIAL).
FT	DOMAIN	377	415		GLY/PRO-RICH.
ET	DOMAIN	476	601		GLY/PRO-RICH.
ET	DOMAIN	610	905		PROTEIN KINASE.
ET	NP_BIND	616	624		ATP (BY SIMILARITY).
ET	BINDING	655	655		ATP (BY SIMILARITY).
ET	ACT SITE	766	766		(BY SIMILARITY).
ET	DISTURF	31	185		BY SIMILARITY.
ET	MOD_RES	513	513		PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
ET	MOD_RES	792	792		PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
ET	MOD_RES	796	796		PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
ET	MOD_RES	797	797		PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
ET	MOD_RES	797	797		PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
ET	CARBONHD	211	211		N-LINKED (GLCNAC. .) (POTENTIAL).
ET	CARBONHD	260	260		N-LINKED (GLCNAC. .) (POTENTIAL).
ET	CARBONHD	371	371		N-LINKED (GLCNAC. .) (POTENTIAL).
ET	CARBONHD	394	394		N-LINKED (GLCNAC. .) (POTENTIAL).
ET	VARSPLIC	506	542		MISSING (IN ISOFORM 2).
ET	CONFLICT	94	94		L -> V (IN REF. 2 AND 3).
ET	CONFLICT	833	833		L -> V (IN REF. 2 AND 3).
ET	CONFLICT	847	867		OUTLEVENNENGEFFDQGRQ -> SAHRRAGHRERGCVLP GCGQA (IN REF. 4).
SQ	SEQUENCE	913 AA;	101127 MM;		C96913EA906C481E CRC64;

```

Query Match Similarity      7.9%; Score 84; DB 1; Length 913;
Best Local Similarity      22.7%; Pred. No. 4.1;
Matches 39; Conservative 20; Mismatches 73; Indels 40; Gaps 7.

OY      7 NOTSPGATEDVSYGSMYIDEPO----GKXELQPEGEVVSCHT-STIPGLYHACLASLSLT 61
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      370 NNSSPALGCTGPPPAWMPWPPEPTNFSLSLEPRGQQPVAKAGSPITALICLVAILLL 429
OY      62 VILLIAMLV-----RRRLMPD-----CVRGRPGLPSEVDFP---- 93
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      430 LLIIITMLMLRMLHMRRLLSKARRVLEELVYHLSVPDITLIIINRRPGREPPVGEPRP 489
OY      94 AGDRPRAVPAVAFVWLLSSLCILLPEDALPPLUTLASAPSDGKTEAPRGAW 145
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      490 RGNPPHSAPC---VNGSALLSLNPAVRLLATVAPRPGPG---PPTPAW 534

RESULT 4
CEBB_RAT
ID      CEBB_RAT      STANDARD;      PRT;      281 AA.
AC      P56261;
DT      15-JUL-1998 (Rel. 36, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      CCAAT/enhancer binding protein epsilon (C/EBP epsilon) (C/EBP-related
      protein 1).
DE      CEBP OR CRP1.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RN      SEQUENCE FROM N.A.
RX      MEDLINE=98256268; PubMed=9593684;
RA      Williams S.C., Du Y., Schwartz R.C., Weiler S.R., Ortiz M.,
RA      Keller J.R., Johnson P.F.;
RA      "/EBPepsilon is a myeloid-specific activator of cytokine, chemokine,
RT      and macrophage-colony-stimulating factor receptor genes."/;
RN      J. Biol. Chem. 273:13493-13501(1998).
RN      [2]
RN      SEQUENCE OF 31-281 FROM N.A.
RX      MEDLINE=91357471; PubMed=1864998;
RA      Williams S.C., Cantwell C.A., Johnson P.F.;
RA      "A family of C/EBP-related proteins capable of forming covalently
RT      linked dimer zipper dimers in vitro.";
RN

```

RL Genes Dev. 5:1553-1567(1991).

CC -1- FUNCTION: C/EBP ARE DNA-BINDING PROTEINS THAT RECOGNIZE TWO
CC DIFFERENT MOTIFS: THE CCAAT HOMOLOGY COMMON TO MANY PROMOTERS AND
CC THE ENHANCED CORE HOMOLOGY COMMON TO MANY ENHANCERS (BY
CC SIMILARITY).

CC -1- SUBUNIT: BINDS DNA AS A DIMER AND CAN FORM STABLE HETERODIMERS
CC WITH C/EBP DELTA.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- SIMILARITY: TO OTHER BZIP PROTEINS. STRONG, TO OTHER C/EBP
CC PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AF034716; AAC24455.1; -.

CC InterPro: IPR004827; TF_BZIP.

CC Pfam: PF00170; bzip, 1.

CC SMART: SM00338; BRZ, 1.

CC Transcription regulation; Activator; DNA-binding; Nuclear protein.

CC DNA_BIND 208 228 BASIC MOTIF.

CC FT DOMAIN 239 267 LEUCINE-ZIPPER.

CC SEQUENCE 281 AA; 30589 MW; 68B16455C034E250 CRC64;

Query Match 7.8%; Score 83; DB 1; Length 281;
Best Local Similarity 28.1%; Pred. No. 1.5;
Matches 50; Conservative 17; Mismatches 65; Indels 46; Gaps 11;

QY 18 SYGSGWIDPEQGXELQP-----EGEV-PSCHTSPGGLYHACLASLILV----- 62
DB 2 SHGTVECEPRGQ--QPLEFSGRAGPELGMC-----HEASIDSAVIESGE 51
QY 63 LLLAMLVRRRLQMPDVCGRPELPVPVPLAGD-RPRAVPAVFWVLSSLCLLPDE 121
DB 52 QLSLDFPANKPPEARSLKNG-PETSPFHYLPADPPFPFYPHRTG-----PPRK 100
QY 122 ALPFLTLASAPSDGKT-----EAPRG--AMKILGLFHYALYYPLAACATAGTAHL 173
DB 101 ALGPGIVSPGSDYDPAVAVKEPRGECNRTGSGSYPLQYVAHC---GQTVHL 155

RESULT 5
NCAP_PHV STANDARD; PRT; 433 AA.
ID NCAP_PHV STANDARD; PRT; 433 AA.
AC P22047;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nucleocapsid protein (Nucleoprotein).
GN S.
OS Prospect Hill virus (PHV).
OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Hantavirus.
OX NCBI_TaxID=11603;
RN [1] TaxID=11603;
RP SEQUENCE FROM N.A.
RA STRAIN=PHV-1;
RA MEDLINE=90177212; PubMed=2309440;
RT "Partridge M.A., Kang C.Y.;
RT "Nucleotide sequence analysis of the S genomic segment of Prospect
RT Hill virus: comparison with the prototype Hantavirus";
RL Virology 175:167-175(1990).
CC -1- SUBCELLULAR LOCATION: INTERNAL PROTEIN OF VIRUS PARTICLE.
CC -1- SIMILARITY: TO OTHER HANTAVIRUS S SEGMENT PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M34011; AAA47086.1; -.

CC PIR: A34681; VHVP.

CC InterPro: IPR002214; Hanta_nucleocap.

CC Pfam: PF00846; Hanta_nucleocap, 1.

CC ProDom: PD001501; Hanta_nucleocap, 1.

CC KM Nucleocapsid; RNA-binding.

CC CONFLICT 7 7 I -> T (IN REF. 1; AAA47086).
CC SEQUENCE 433 AA; 49005 MW; EBBDB161AD32A0CD CRC64;

Query Match 7.5%; Score 80; DB 1; Length 433;
Best Local Similarity 24.1%; Pred. No. 4.4;
Matches 35; Conservative 21; Mismatches 55; Indels 34; Gaps 6;

QY 23 YIDPEQGXELQPEGEVPSCHTSPGGLYHACLASLILVLL-----LAMVRRRLQMPD 78
DB 178 YVSMPTAGSTWKAELTGRFRITVCGLPAGQIMARNIISPMGVIGFAFVKD---WAD 234
QY 79 CYRG-----RPGGPS-PVDFLAGDRPRAVPAVFWVLSSLCLL-----LPDE 120
DB 235 KYKAFLDGKCPFLAEPKPGQPAGEAEFLSIR-----AVIMNRQAVLDETHLPDI 285
QY 121 DALPFLTLASAPSDGKTAPRGAW 145
DB 286 DALVEILASGDPITPDSLLENPHAM 310

RESULT 6
NFC2_HUMAN STANDARD; PRT; 925 AA.
ID NFC2_HUMAN STANDARD; PRT; 925 AA.
AC Q13469; Q13468; Q9NRY6; Q9NRY3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nuclear factor of activated T-cells, cytoplasmic 2 (T cell
DE transcription factor NFAT1) (NFAT pre-existing subunit) (NF-ATp).
GN NFATC2 OR NFAT1 OR NFATP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=96251346; PubMed=8668213;
RA Luo C., Burgeon E., Carew J.A., McCaffrey P.G., Badalian T.M.,
RA Lane W.S., Hogan P.G., Rao A.;
RT "Recombinant NFAT1 (NFATp) is regulated by calcineurin in T cells and
RT mediates transcription of several cytokine genes";
RL Mol. Cell. Biol. 16:3955-3966(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A.G., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levenshaji M.H., Leverhwa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McIlroy K., Murray A.A.,
RA Miles S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prichardlingam S.R., Plumb R.W., Ramsey H.,
RA Rice C.M., Rose M.T., Scott C.E., Sehara H.K., Shownkeen R., Sims S.,

RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall J.M., Wallis J.M., Williams S.A.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 "The DNA sequence and comparative analysis of human chromosome 20.";
 Nature 414:865-871(2001).
 [3]
 RN Nature 414:865-871(2001).
 RP REVIEW.
 RX MEDLINE=99189746; PubMed=10089876;
 RA Crabtree G.R.;
 RT "Genetic signals and specific outcomes: signaling through Ca2+,"
 RT calcineurin, and NF-AT.";
 RL Cell 96:611-614(1999).
 CC -1- FUNCTION: PLAYS A ROLE IN THE INDUCIBLE EXPRESSION OF CYTOKINE
 CC GENES IN T CELLS, ESPECIALLY IN THE INDUCTION OF THE IL-2, IL-3,
 CC IL-4, TNF-ALPHA OR GM-CSF.
 CC -1- SUBUNIT: MEMBER OF THE MULTICOMPONENT NFATC TRANSCRIPTION COMPLEX
 CC THAT CONSISTS OF AT LEAST TWO COMPONENTS, A PRE-EXISTING
 CC CYTOPLASMIC COMPONENT NFATC2 AND AN INDUCIBLE NUCLEAR COMPONENT
 CC NFATC1. OTHER MEMBERS SUCH AS NFATC4, NFATC3 OR MEMBERS OF THE
 CC ACTIVATING PROTEIN-1 FAMILY, MAF, GATA4 AND CBP/300 CAN ALSO BIND
 CC THE COMPLEX. NFATC PROTEINS BIND TO DNA AS MONOMERS.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FOR THE PHOSPHORYLATED FORM AND
 CC NUCLEAR AFTER ACTIVATION THAT IS CONTROLLED BY CALCINEURIN-
 CC MEDIATED DEPHOSPHORYLATION. RAPID NUCLEAR EXIT OF NFATC IS THOUGHT
 CC TO BE ONE MECHANISM BY WHICH CELLS DISTINGUISH BETWEEN SUSTAINED
 CC AND TRANSIENT CALCIUM SIGNALS. THE SUBCELLULAR LOCALIZATION OF
 CC NFATC PLAY A KEY ROLE IN THE GENE TRANSCRIPTION.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; B AND C (SHOWN HERE);
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, SPLEEN, HEART, TESTIS,
 CC BRAIN, PLACENTA, MOSCLE AND PANCREAS.
 CC -1- INDUCTION: INDUCIBLY EXPRESSED IN T LYMPHOCYTES UPON ACTIVATION OF
 CC THE T-CELL RECEPTOR (TCR) COMPLEX. INDUCED AFTER CO-ADDITION OF
 CC PHORBOL 12-MYRISTATE 13-ACETATE (PMA) AND IONOMYCIN.
 CC -1- DOMAIN: REL SIMILARITY DOMAIN (RSD) ALLOWS DNA-BINDING AND
 CC COOPERATIVE INTERACTIONS WITH AP1 FACTORS (BY SIMILARITY).
 CC -1- PTM: PHOSPHORYLATED BY NFATC-KINASE; DEPHOSPHORYLATED BY
 CC CALCINEURIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U43341; AAC50886.1; -
 CC EMBL; U43342; AAC50887.1; -
 CC EMBL; AL035682; CAC00529.1; -
 CC EMBL; AL035682; CAB54871.1; -
 CC EMBL; AL132866; CAC00528.1; -
 CC TRANSFAC; T01948; -
 CC Gene; HGNC:7776; NFATC2.
 DR MIM; 600490; -
 DR InterPro; IPR002909; IPT_TIG.
 DR InterPro; IPR000451; NF_Rel_dor_fam.
 DR Pfam; PF00554; RHD; 1.
 DR Pfam; PF01833; TIG; 1.
 DR SMART; SM00429; IPT; 1.
 DR PROSITE; PS01204; REL_1; FALSE_NEG.
 DR PROSITE; PS50254; REL_2; 1.
 KM Transcription regulation; Activator; Nuclear protein; DNA-binding;
 KM Alternative splicing; Phosphorylation; Repeat.
 FT DOMAIN 111 116
 FT 119 199 TRANS-ACTIVATION DOMAIN A (TAD-A).
 FT DOMAIN 184 286
 FT REPEAT 184 200
 FT REPEAT 213 229
 FT 2.

FT REPEAT 272 286 3 (APPROXIMATE).
 FT DOMAIN 251 253 NUCLEAR LOCALIZATION SIGNAL.
 FT DOMAIN 421 428 DNA-BINDING.
 FT DOMAIN 664 666 NUCLEAR LOCALIZATION SIGNAL.
 FT DOMAIN 904 913 NUCLEAR EXPORT SIGNAL.
 FT MOD RES 110 110 PHOSPHORYLATION (BY SIMILARITY).
 FT VARSPLIC 908 925 VNETIRKFSGPPARKNOT -> ELIDTHLSMIGNIL (IN ISOFORM B).
 FT CONFLICT 65 65 M -> L (IN REF. 2).
 SQ SEQUENCE 925 AA; 100163 MW; A825164A9C279025 CRC64;
 Query Match 7.54; Score 80; DB-1; Length 925;
 Best Local Similarity 23.54; Pred. No. 9.7;
 Matches 57; Conservative 30; Mismatches 76; Indels 80; Gaps 15;
 QY 3 OPAGNOTSPGATEDYSDYSWYIDEPQGX-----ELQPEGEV-PSCHTSPGLYHAC 54
 DB 240 RPSRSSSPGAKRRHSACAEALVALPFGASPQRSRSPQSPSHVAPQHG--YPP 296
 QY 55 LASLTLVLLLLAMLVRR-----RLW-----PDCVGRNP---GLP---SPYRLA--- 94
 DB 297 VAGSAVIMDALNSLATDSPCGIPPKMWTSPSPVSAAPSXAGLPRIHYPAVEFLAPCE 356
 QY 95 -GDRPAVPAAPFVWLLSLCLLDDEALPPLT---ASAP-----SQDG----- 136
 DB 357 QGRRNSAPESILVPPTPKPLVP---ALPICSIPVTRSLPPLMPLSSQSGSYELRIE 413
 QY 137 -----KTEAPRGAMKILGLFHYAALYPLAACATAGTAHLIGSTLSMAHLGV 185
 DB 414 VQKPHRHRAHYETSGRAVK-----APTGGHPVQLHG-VMEKPKPLD 456
 QY 186 QVW 188
 DB 457 QIF 459
 RESULT 7
 DDRI_RAT STANDARD; PRT; 910 AA.
 AC 063474;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epithelial discoidin domain receptor 1 precursor (BC 2.7.1.112)
 DE (Tyrosine-protein kinase CAK) (cell adhesion kinase) (Tyrosine kinase
 DE DDR) (Discoidin receptor tyrosine kinase) (Protein-tyrosine kinase
 DE PTK-3).
 GN DDRI OR EDDR1 OR PTK3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=94173920; PubMed=8127887;
 RA Sanchez M.P., Tapley P., Saint S.S., He B., Pulido D., Barbacid M.;
 RT "Multiple tyrosine protein kinases in rat hippocampal neurons:
 RT isolation of Ptk-3, a receptor expressed in proliferative zones of
 RT the developing brain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:1819-1823(1994).
 CC -1- FUNCTION: MAY BE INVOLVED IN CELL-CELL INTERACTIONS AND
 CC RECOGNITION (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: VARIOUS EMBRYONIC AND ADULT TISSUES; ALSO
 CC PROLIFERATIVE ZONES OF THE DEVELOPING BRAIN; HIPPOCAMPAL NEURONS.
 CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF TYR-PROTEIN
 CC KINASES.
 CC -1- SIMILARITY: CONTAINS 1 FS/8 TYPE C DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
 CC RECEPTOR SUBFAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).

CC -----
 CC EMBL: L26525; AAA21089.1; -
 CC DR HSSP; P00523; 2PTK.
 CC DR InterPro: IPR000719; Burk kinase.
 CC DR InterPro: IPR000423; PAK5_C.
 CC DR InterPro: IPR002011; RtkKinaseII.
 CC DR InterPro: IPR001245; Tyr_kinase.
 CC DR Pfam; PF00754; Pk_kinase; 1.
 CC DR Pfam; PF00754; P5_P8 type C; 1.
 CC DR ProDom; PD000001; Burk_kinase; 1.
 CC DR SMART; SM00231; PAK5C; 1.
 CC DR SMART; SM00219; Tyrc; 1.
 CC DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 CC DR PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
 CC DR PROSITE; PS01285; PAK5C_1; 1.
 CC DR PROSITE; PS01286; PAK5C_2; 1.
 CC DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 CC DR Transferrase; Tyrosine-protein kinase; Glycoprotein; Signal;
 CC KM Phosphorylation; Transmembrane; Receptor; ATP-binding.
 CC FT SIGNAL 1 19
 CC FT CHAIN 20 910
 CC FT DOMAIN 20 413
 CC FT TRANSMEM 414 440
 CC FT DOMAIN 441 910
 CC FT DOMAIN 32 186
 CC FT DOMAIN 378 412
 CC FT DOMAIN 473 538
 CC FT NP_BIND 607 902
 CC FT BINDING 613 621
 CC FT ACT_SITE 652 652
 CC FT ACT_SITE 763 763
 CC FT DISULFID 32 186
 CC FT MOD_RES 510 510
 CC FT MOD_RES 789 789
 CC FT MOD_RES 793 793
 CC FT MOD_RES 794 794
 CC FT CARBOHYD 212 212
 CC FT CARBOHYD 261 261
 CC FT CARBOHYD 371 371
 CC FT CARBOHYD 391 391
 CC FT SEQUENCE 910 AA; 101164 MW; 757FPAIDCB029806 CRC64;

Query Match 7.4%; Score 79; DB 1; Length 910;
 Best Local Similarity 22.1%; Pred. No. 12;
 Matches 36; Conservative 23; Mismatches 70; Indels 34; Gaps 7;

QY 13 ATEDYSVSYGWIIDEPO---GGKELQPEGEVPSCHT-SIPGLVYHACIASIVLILLIA 67
 DB 373 SSDTPPAWMPGPPTNPFSSLEPRGQGVAVAKESPAALIGCVAILLILLIA 432
 QY 68 MLYV-----RRQLWMD-----CYRGPRGLSPVDVFLAGDPPRAVPA 103
 DB 433 LMLMRLHRRLLSKAEKRVLEELTVHLSPQDITLLINNRGQPREPPPY-GEPRPRGTP 491
 QY 104 AVFWLLSLCLLPDED-ALPFLTLGAPSDQKTEAPRAAW 145
 DB 492 HSAPCVFPGSALLLSNPAYRLLATYAAPPGPG---PPTPAW 531

RESULT 8
 GRB7_HUMAN STANDARD; PRT; 532 AA.
 AC Q1451; Q92568; Q96DF9.
 DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Growth factor receptor-bound protein 7 (GRB7 adapter protein)
 DE (Epidermal growth factor receptor GRB-7) (B47).
 GN GRB7.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RX MEDLINE=97236270; PubMed=9125150;
 RA Kiani T., Saeki H., Akiyama N., Ishizuka T., Sakamoto H., Aizawa S.,
 RA Sugimachi T., Terada M.;
 RT "Molecular cloning of human GRB-7 co-amplified with CABI and c-ERBB-2
 RT in primary gastric cancer";
 RL Biochem. Biophys. Res. Commun. 232:5-9(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98376491; PubMed=9710451;
 RA Tanaka S., Mori M., Akiyoshi T., Tanaka Y., Mafune K., Wands J.R.,
 RA Sugimachi K.;
 RT "A novel variant of human Grb7 is associated with invasive esophageal
 RT carcinoma.";
 RL J. Clin. Invest. 102:821-827(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Whitlock N.V., Sady R.A.J., McGrath J.A.;
 RT "Genomic organization and amplification of the human GRB7 gene";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strausberg R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 130-343 FROM N.A.
 RX MEDLINE=97141776; PubMed=8988034;
 RA Tanaka S., Mori M., Akiyoshi T., Tanaka Y., Mafune K., Wands J.R.,
 RA Sugimachi K.;
 RT "Overexpression of Grb7 with epidermal growth factor receptor or
 RT Her2/erbB2 in human advanced esophageal carcinoma";
 RL Cancer Res. 57:28-31(1997).
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC DOMAIN OF THE EPIDERMAL
 CC GROWTH FACTOR RECEPTOR WHICH IS THEN INHIBITED. THE INTERACTION IS
 CC MEDIATED BY THE SH2 DOMAIN. ALSO BINDS TO ERBB2.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING.
 CC -1- PTM: PHOSPHORYLATED ON TYROSINE RESIDUES.
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE GRB7/10/14 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).

CC -----
 CC EMBL: D43772; BAA07827.1; -
 CC DR EMBL; AB008789; BAA29059.1; -
 CC DR EMBL; AF274875; AAG25938.1; -
 CC DR EMBL; BC006535; AAH06535.1; -
 CC DR EMBL; D87513; BAA13412.1; -
 CC DR HSSP; P00519; IAB2.
 CC DR Genew; HGNC:4567; GRB7.
 CC DR MIM; 601522; -
 CC DR InterPro; IPR001849; PH.
 CC DR InterPro; IPR000159; RA domain.
 CC DR InterPro; IPR000980; SH2.

DR Pfam: PF00017; SH2; 1.
 DR Pfam: PF00159; PH; 1.
 DR Pfam: PF00788; RA; 1.
 DR PRINTS; PR00401; SH2DOMAIN.
 DR PRODOM; PD000093; SH2; 1.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00314; RA; 1.
 DR SMART; SM00252; SH2; 1.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 DR PROSITE; PS50001; SH2; 1.
 DR SH2 domain; Phosphorylation; Alternative splicing.
 FT DOMAIN 229 338 PH.
 FT DOMAIN 431 512 SH2.
 FT CONFLICT 18 18 W -> C (IN REF. 4).
 SQ SEQUENCE 532 AA; 59764 MW; A68679F83A146F74 CRC64;

Query Match 7.4%; Score 78.5; DB 1; Length 532;
 Best Local Similarity 28.3%; Pred. No. 7.4;
 Matches 39; Conservative 12; Mismatches 62; Indels 25; Gaps 7;

QY 75 LMPDCRCRGLPSVDF-LGDRRAVPAVFMVLLSLCLLPDEDA-----P 124
 DB 17 LWP-APGTPGTPPPPTDPLPEVRSQP-----LLPTGRKREBRRTSLPSIPNP 70
 QY 125 FLTLASAPSDGKTEAPRGAWKIL---GLPHYALYYPLAAC---ATAGHTAHLIGS 176
 DB 71 FPELCSPSSPILIGPSASAGLPRDASRPHVVKVISEDGACRSVEVAAGATARHCEM 130
 QY 177 TLSSAH-LGVVWORAEC 193
 DB 131 LVGRRAHLSDETGLVEEC 148

RESULT 9
 CDS OEENBE STANDARD; PRT; 577 AA.
 AC 004648;
 ID 01-JUN-1994 (Rel. 29, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Probable cytochrome c biosynthesis protein.
 OS Oenothera lamarckiana (Bertero's evening primrose).
 OC Mitochondrion.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Myrtales; Onagraceae; Oenothera.
 OC NCBI TaxId=3950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93288005; PubMed=8389979;
 RA Schuster W., Combes B., Pileger K., Brennicke A.;
 RA "A plant mitochondrial gene encodes a protein involved in cytochrome
 c biogenesis";
 RT Mol. Gen. Genet. 239:49-57(1993).
 RL -1- FUNCTION: COULD BE INVOLVED IN ASSEMBLY AND MATURATION OF
 CC CYTOCHROMES C. MAY PLAY A ROLE IN GUIDANCE OF APOCYTOCHROMES AND
 CC HEME GROUPS FOR THE COVALENT LINKAGE INTRODUCED BY THE CYTOCHROME-
 CC C-HEME LYASE.
 CC -1- SIMILARITY: BELONGS TO THE CCMF/CYCK/CELI/NRE/CESA FAMILY.
 CC -1- CAUTION: EXTENSIVELY MODIFIED BY RNA-EDITING. WE HAVE ONLY
 CC MODIFIED THE GENOMIC SEQUENCE FOR THE OBSERVED R-W EDITING.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X69555; CAA9287.1; ALT_SEQ.
 DR InterPro; IPR002541; CytC_asm.
 DR InterPro; IPR003567; Cyt_c_bilog.

DR Pfam: PF01578; CytC_asm; 2.
 DR PRINTS; PR01410; CCBIOGENESIS.
 KW Mitochondrion; Cytochrome c-type biogenesis; RNA editing.
 SQ SEQUENCE 577 AA; 65238 MW; 10FB0F97228B8AB8 CRC64;

Query Match 7.4%; Score 78.5; DB 1; Length 577;
 Best Local Similarity 24.5%; Pred. No. 8.1;
 Matches 27; Conservative 18; Mismatches 46; Indels 19; Gaps 4;

QY 13 ATEDVSGSMYIDEPQGXELQPEG-EVPSCHTSIPPGIYHACLASITLVLLAMVLR 71
 DB 438 AHHELGMGGMWMDPPEVNSFMPRVATARIHSVILP-LLHSTSLINVTLLC----- 490
 QY 72 RROLWPDVYRG---RPGLPSPVDFLAGDRPRAVPAVFMVLLSLCLLP 117
 DB 491 -----CVLGTSIRSGLLASVHSFATDTRGIFLMRPFLLMTGISMTL 533

RESULT 10
 OL56 STRAT STANDARD; PRT; 3519 AA.
 ID 007017;
 AC 007017;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Oleandomycin polyketide synthase, modules 5 and 6.
 DE ORF.
 OS Streptomyces antibioticus.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OC NCBI_TaxId=1890;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150470; PubMed=8107683;
 RA Swan D.G., Rodriguez A.M., Vilches C., Salas J.A.;
 RA "Characterization of a Streptomyces antibiotic gene encoding a type
 RT I polyketide synthase which has an unusual coding sequence."
 RL Mol. Gen. Genet. 242:358-362(1994).
 CC -1- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
 CC LACTONE RING.
 CC -1- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
 CC -1- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L09654; AAA19695.1; --
 DR HSPB; P25715; IMLA.
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR000794; ketoacyl-synth.
 DR InterPro; IPR003880; pantoic_acid.
 DR InterPro; IPR001031; Thioesterase.
 DR Pfam; PF00109; ketoacyl-synth; 2.
 DR Pfam; PF00550; pp-binding; 2.
 DR Pfam; PF00698; Acyl-transf; 2.
 DR Pfam; PF00975; Thioesterase; 1.
 DR Pfam; PF02801; ketoacyl-synth; 2.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
 DR PROSITE; PS00606; B-KETOACYL SYNTHASE; 2.
 DR PROSITE; PS00075; ACP DOMAIN; 2.
 DR TRANSFERASE; Acyltransferase; Antibiotic biosynthesis; NADP;
 DR phosphopantetheine; Multifunctional enzyme; Repeat.
 FT DOMAIN 1 3519 MODULE 5.
 FT DOMAIN ? 501 MODULE 6.
 FT DOMAIN 32 501 BETA-KETOACYL SYNTHASE 1.
 FT DOMAIN 569 890 ACYLTRANSFERASE (AT) 1.
 FT DOMAIN 1200 1382 BETA-KETOACYL REDUCTASE 1.
 FT DOMAIN 1487 1561 ACYL CARRIER (ACP) 1.

FT DOMAIN 1686 2156 BETA-KETOACYL SYNTHASE 2.
 FT DOMAIN 2220 2541 ACYLTRANSFERRASE (AT) 2.
 FT DOMAIN 2856 3038 BETA-KETOACYL REDUCTASE 2.
 FT DOMAIN 3141 3215 ACYL CARRIER (ACP) 2.
 FT ACT SITE ? ? ACTL-ENZYMATIC INTERMEDIATE.
 FT BINDING ? ? PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT DOMAIN 3270 3519 THIOESTERASE.
 FT ACT SITE 210 210 ACYL-KETOACYL SYNTHASE.
 FT NP BIND 660 660 ACYL-ENZYMATIC INTERMEDIATE.
 FT BINDING 1203 1249 NADP.
 FT ACT SITE 1524 1524 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT ACT SITE 1859 1859 BETA-KETOACYL SYNTHASE.
 FT ACT SITE 2311 2311 ACYL-ENZYMATIC INTERMEDIATE.
 FT NP BIND 2859 2905 NADP.
 FT BINDING 3178 3178 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 SQ SEQUENCE 3519 AA; 368561 MW; 41AE7BAABE9B1F6 CRC64;

Query Match 7.4%; Score 78.5; DB 1; Length 3519;
 Best Local Similarity 23.5%; Pred. No. 54;
 Matches 54; Conservative 25; Mismatches 70; Indels 81; Gaps 12;

QY 11 PCATEDVSGWYIDPEOGXELDPGEVPSCHTSIPGLYHACLASLIVLLIAMLV 70
 DB 2642 PALDDEVLCGGWLFVVRG-----LADGVAVARVAATA 2676
 QY 71 R-----RRQLWPCVGRPGLPSPVDFLAGDRPR-----AVPA--AVFVVL 109
 DB 2677 RGEVSVVELDPTPRPRRAVAEAVAGR-GVSGVVSFLSMDDRHSHPVPPAGLASLVL 2735
 QY 110 LSLIC-----LLRLDEDLPLTLASAPSDCKTEARGA--W--KILGLFH-- 152
 DB 2736 AQAULVGLGVSGPRLMLVTRDA---VVAGPSDAGAVIDPQAVQWVGFGVLGHEPE 2790
 QY 153 --YAAALYPLAA---CATAGHTAAHLGSLTSLVAHL-----GVQVMOR 190
 DB 2791 LMGGLIDLPEVPEPESGTCDHRYADLATVVASAGFEDVAVRSGSVWR 2840

RESULT 11

CEBE_HUMAN STANDARD; PRT; 281 AA.
 ID 15744; Q15745; Q09803;
 AC Q15744; Q15745; Q09803; Created
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE CCAT/enhancer binding protein epsilon (C/EBP epsilon).
 GN CEBPE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=96299737; PubMed=8661101;
 RA Antonson P., Stellan B., Yamanaka R., Xanthopoulos K.G.;
 RT "A novel human CCAAT/enhancer binding protein gene, C/EBPepsilon, is
 RT expressed in cells of lymphoid and myeloid lineages and is localized
 RT on chromosome 1q11.2 close to the T-cell receptor alpha/delta
 RT locus";
 RT Genomics 35:30-38(1996).
 RL [2]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RP MEDLINE=97184462; PubMed=9032264;
 RA Chumakov A.M., Gyller I., Chumakova E., Chih D., Slater J.,
 RA Koefler H.P.;
 RT "Cloning of the novel human myeloid-cell-specific C/EBP-epsilon
 RT transcription factor";
 RT Mol. Cell. Biol. 17:1375-1386(1997).
 RL [1]
 RP FUNCTION: C/EBP ARE DNA-BINDING PROTEINS THAT RECOGNIZE TWO
 RP DIFFERENT MOTIFS: THE CCAAT HOMOLOG COMMON TO MANY PROMOTERS AND
 RP THE ENHANCED CORE HOMOLOG COMMON TO MANY ENHANCERS.
 CC -1 SUBUNIT: BINDS DNA AS A DIMER AND CAN FORM STABLE HETERODIMERS
 CC WITH C/EBP DELTA.

CC -1 SUBCELLULAR LOCATION: Nucleus.
 CC -1 TISSUE SPECIFICITY: Strongest expression occurs in promyelocyte
 CC and late-myeloblast-like cell lines.
 CC -1 PTM: Phosphorylated.
 CC -1 SIMILARITY: TO OTHER BZIP PROTEINS. STRONG, TO OTHER C/EBP
 CC PROTEINS.
 CC -1 CAUTION: Ref.2 sequence differs from that shown due to a
 CC framehift in position 4.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U48865; AAC50708.1; -;
 DR EMBL; U48866; AAC50709.1; -;
 DR EMBL; U80982; AAC51130.1; ALT_FRAME.
 DR TRANSFAC; T04883; -;
 DR GeneW; HGNC:1836; CEBPE.
 DR MIM; 600749; -;
 DR InterPro; IPR004827; TF_BZIP.
 DR Pfam; PF00170; bZIP; 2.
 DR SMART; SM00338; BRLZ; 1.
 KW Transcription regulation; Activator; DNA-binding; Nuclear protein;
 KW phosphorylation.
 FT DNA_BIND 208 228 BASIC MOTIF.
 FT DOMAIN 239 267 LEUCINE-ZIPPER.
 FT ACT 64 64 P -> S (IN REF. 2).
 FT CONFLICT 68 68 R -> G (IN REF. 2).
 FT CONFLICT 252 252 Q -> E (IN REF. 2).
 SQ SEQUENCE 281 AA; 30702 MW; 2F808717AD5ACD2B CRC64;

Query Match 7.4%; Score 78; DB 1; Length 281;
 Best Local Similarity 28.9%; Pred. No. 4.2;
 Matches 52; Conservative 16; Mismatches 62; Indels 50; Gaps 11;

QY 18 SYGWSYIDPEOGXELDPGEVPSCHTSIPGLYHACLAS-----LST 60
 DB 2 SHGTYVCEPQGQ--OPLFSGRAGPGLDMCEHSAIDLSAY---IESGEQLLSD 56
 QY 61 LVLLDLMLVRRRLMDCCVRGPGLPSPVDFLAGD-RPAVPAVPMVLLSLCLLLP 119
 DB 57 LFAVKRPAEARR-----LKG-PGTPAPHYLPDPDRPAPVPPHTFG-----PD 98
 QY 120 EDALPFLTLASAPSDCKT---EAPRG--AWKILGLFYAALYPLAACATAGHTAAHL 173
 DB 99 RKALGPPIVSPGSDYDRAVAVAKKEPRGEGSRAARGSYNPLQYVANC---GOTAMHL 155

RESULT 12

PFTB_PEA STANDARD; PRT; 419 AA.
 ID Q04903;
 AC Q04903;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein farnesyltransferase beta subunit (EC 2.5.1.-) (CAAX
 DE farnesyltransferase beta subunit) (RAS proteins prenyltransferase
 DE beta) (Ftase-beta).
 GN PFTB.
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC euroside 1; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
 OC NCB1_TaxID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=cv. Alaska; TISSUE=Seedling;
 RX MEDLINE=94105305; PubMed=8278509;
 RA Yang Z., Cramer C.L., Watson J.C.;


```

FT DOMAIN 442 911 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 32 186 F5/8 TYPE C (PHOSPHOLIPID-BINDING,
FT 379 413 POTENTIAL).
FT DOMAIN 474 599 GLY/PRO-RICH.
FT DOMAIN 608 903 GLY/PRO-RICH.
FT NP_BIND 614 622 ATP (BY SIMILARITY).
FT BINDING 653 653 ATP (BY SIMILARITY).
FT ACT_SITE 764 764 BY SIMILARITY.
FT DISULFID 32 186 BY SIMILARITY.
FT MOD_RES 511 511 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 750 750 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 794 794 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 795 795 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VAASPLIC 503 539 MISSING (IN ISOFORM CAK 11).
SQ SEQUENCE 911 AA; 101160 MW; DBB7FE03DD79510 CRC64;

```

```

Query Match 7.4%; Score 78; DB 1; Length 911;
Best Local Similarity 22.1%; Pred. No. 14;
Matches 36; Conservative 23; Mismatches 70; Indels 34; Gaps 7;

```

```

OY 13 ATEDYSGWYIDEPQ---GGXELQPEGEVPSCHT-SIPGLYHACLASISLVLLLLA 67
DB 374 SDDTPPAWMPWPFPPTNFSSLELEPRGQPVAKAEGSPTILILGCVAILLLILLIA 433
OY 68 MLV-----RRLOMPD-----CVGRGRLPPVPVDFLACDRPRAVPA 103
DB 434 LMLWRLHWRRLLSKARERLEELTVHLSVPGDTLLINRRPBRPPPY-QEPRRGCTPP 492
OY 104 AVFVLLSLCLLPDED-ALPFLTLASAPSDGKTEAPRGAM 145
DB 493 HSAPCPVPGSALLLSNPVRLILATYARPPPGP---PPTPAM 532

```

```

RESULT 14
CCBS_DAUCA STANDARD; PRT: 579 AA.
AC 00467;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Probable cytochrome c biosynthesis protein.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
ON NCBI_TaxID=4039;
RX MEDLINE=93288005; Pubmed=8369979;
RA Schuster W., Combettes B., Flieger K., Brennicke A.;
RT "A plant mitochondrial gene encodes a protein involved in cytochrome
RT c biogenesis.";
RL Mol. Gen. Genet. 239:49-57(1993).
CC -1- FUNCTION: COULD BE INVOLVED IN ASSEMBLY AND MATURATION OF
CC CYTOCHROMES C. MAY PLAY A ROLE IN GUIDANCE OF APOCYTOCHROMES AND
CC HEME GROUPS FOR THE COVALENT LINKAGE INTRODUCED BY THE CYTOCHROME-
CC C-HEME LYASE.
CC -1- CAUTION: BELONGS TO THE CCMF/CYCK/CCIL/NRPE/CCSA FAMILY.
CC -1- CAUTION: EXTENSIVELY MODIFIED BY RNA-EDITING.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).

```

```

CC EMBL: X69554; CAA43286.1; -
DR PIR: S31568; S31568.
DR InterPro: IPR002541; Cyto_c_asm.
DR InterPro: IPR003567; Cyto_c_biol.
DR Pfam: PF01578; Cyto_c_asm. 2.
DR PRINTS: PRO1410; CCBIOGENESIS.
KW Mitochondrion; Cytochrome c-type biogenesis; RNA editing.
SQ SEQUENCE 579 AA; 65691 MW; DB2PFC2AFB12F37 CRC64;

```

```

Query Match 7.3%; Score 77.5; DB 1; Length 579;
Best Local Similarity 25.5%; Pred. No. 10;
Matches 28; Conservative 16; Mismatches 47; Indels 19; Gaps 4;

```

```

OY 13 ATEDYSGWYIDEPQGGXELQPEGEVPSCHTSIPGLYHACLASISLVLLLLMLVR 71
DB 440 AHHELGRGGMWFRPDVENASFPVLATARIHSVILP-LHSSWTSFNTIVTL----- 490
OY 72 RROLMPDCVRG---RPLPSPVDFLACDRPRAVPAVFWVLLSLCLL 117
DB 491 -----PCCVSGTSSIRSGLLAPVHSFATDTRGIFLWRFULMIGISMIL 535

```

```

RESULT 15
PASI_MOUSE STANDARD; PRT: 874 AA.
AC P97481; 008787; 055046;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endothelial PAS domain protein 1 (EPAS1), a transcription factor
DE (HLRF) (HIF-related factor) (HRF).
GN EPAS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX MEDLINE=97152468; Pubmed=9000051;
RA Tian H., McKnight S.L., Russell D.W.;
RT "Endothelial PAS domain protein 1 (EPAS1), a transcription factor
RT selectively expressed in endothelial cells.";
RL Genes Dev. 11:72-82(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Hypothalamus, and Skeletal muscle;
RX MEDLINE=97272213; Pubmed=9113979;
RA Ena M., Taya S., Yokotani N., Sogawa K., Matsuda Y.;
RA Fujii-Kuriyama Y.;
RT "A novel bHLH-PAS factor with close sequence similarity to hypoxia-
RT inducible factor 1alpha regulates the VEGF expression and is
RT potentially involved in lung and vascular development.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:4273-4278(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain capillary;
RX MEDLINE=97321546; Pubmed=9178256;
RA Flame T., Froehlich T., von Reutern M., Kappel A., Damert A.;
RA Rissau W.;
RT "HRF, a putative basic helix-loop-helix-PAS-domain transcription
RT factor is closely related to hypoxia-inducible factor-1 alpha and
RT developmentally expressed in blood vessels.";
RL Mech. Dev. 63:51-60(1997).
CC -1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN THE INDUCTION OF OXYGEN
CC REGULATED GENES. SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE
CC ELEMENT (HRE). REGULATES THE VASCULAR ENDOTHELIAL GROWTH FACTOR
CC (VEGF) EXPRESSION AND SEEMS TO BE IMPLICATED IN THE DEVELOPMENT OF
CC BLOOD VESSELS AND THE TUBULAR SYSTEM OF LUNG. MAY ALSO PLAY A ROLE
CC IN THE FORMATION OF THE ENDOTHELIUM GIVING RISE TO THE BLOOD BRAIN
CC BARRIER. POTENT ACTIVATOR OF THE TIE-2 TYROSINE KINASE EXPRESSION.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER

```

BHLH PROTEIN, HETERODIMER WITH THE ARNT PROTEIN.
CC -1- CELLULAR LOCATION: Nuclear (potential)
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES, WITH HIGHEST LEVELS
CC IN LUNG, FOLLOWED BY HEART, KIDNEY, BRAIN AND LIVER. PREDOMINANTLY
CC EXPRESSED IN ENDOTHELIAL CELLS. ALSO FOUND IN SMOOTH MUSCLE CELLS
CC OF THE UTERUS, NEURONS, AND BROWN ADIPOSE TISSUE. HIGH EXPRESSION
CC IN EMBRYONIC CHOROID PLEXUS AND KIDNEY GLOMERULI.
CC -1- DEVELOPMENTAL STAGE: IN DAY 11 EMBryo, EXPRESSION IS ALMOST
CC EXCLUSIVELY SEEN IN ENDOTHELIAL CELLS OF THE INTERSEGMENTAL BLOOD
CC VESSELS SEPARATING THE SOMITES, THE ATRIAL AND VENTRICULAR
CC CHAMBERS OF THE HEART, AND THE DORSAL AORTA. HIGH EXPRESSION ALSO
CC OCCURS IN EXTRA-EMBRYONIC MEMBRANES. IN THE DEVELOPING BRAIN OF
CC DAY 13 EMBryo, ENDOTHELIAL CELLS OF THE HIGHLY VASCULARIZED
CC CHOROID PLEXUS CONTAIN HIGH LEVELS OF EPAS1.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. HIGHEST TO HIF-1 ALPHA.
CC -1- SIMILARITY: CONTAINS 2 PAS (PEP-ARN-T-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.ch/announce/>
CC or send an email to license@isb.ch).
CC -----
DR EMBL; U81983; AAB41496.1; -;
DR EMBL; D89787; BAA20130.1; -;
DR EMBL; AF045160; AAC12871.1; -;
DR TRANSFAC; T02719; -;
DR MGD; MG1:109169; Epas1.
DR InterPro; IPR001092; HLH_Basic.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00765; PAC; 1.
DR Pfam; PF00989; PAS; 2.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 2.
DR TIGRFAMs; TIGR00229; sensory_box; 2.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS01112; PAS; 2.
KW Repeat; DNA-binding; Nuclear protein; Transcription regulation;
KW Activator; Angiogenesis; Developmental protein.
FT FT DNA_BIND 15 27 BASIC DOMAIN.
FT FT DOMAIN 28 68 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT FT DOMAIN 84 154 PAS 1.
FT FT DOMAIN 230 300 PAS 2.
FT FT DOMAIN 304 347 PAC.
FT FT DOMAIN 471 479 POLY-SER.
FT FT CONFLICT 25 25 C -> S (IN REF. 2).
FT FT CONFLICT 191 191 K -> KS (IN REF. 1).
FT FT CONFLICT 439 440 VS -> AA (IN REF. 3).
FT FT CONFLICT 463 463 D -> G (IN REF. 3).
FT FT CONFLICT 654 654 G -> V (IN REF. 2).
FT FT CONFLICT 663 663 A -> P (IN REF. 2).
FT FT CONFLICT 669 669 S -> W (IN REF. 1).
FT FT CONFLICT 673 673 P -> L (IN REF. 1).
FT FT CONFLICT 678 678 P -> L (IN REF. 1).
FT FT CONFLICT 725 725 D -> E (IN REF. 3).
FT FT CONFLICT 731 731 P -> L (IN REF. 3).
FT FT CONFLICT 762 762 A -> G (IN REF. 3).
FT FT CONFLICT 786 786 P -> L (IN REF. 3).
FT FT CONFLICT 791 791 S -> F (IN REF. 3).
FT FT CONFLICT 794 794 S -> N (IN REF. 3).
SQ SEQUENCE 874 AA; 96712 MW; A6FFFA90AB43640C CRC64;

Query Match 7.2%; Score 76.5; DB 1; Length 874;
Best local similarity 22.9%; Pred No. 19;
Matches 61; Conservative 17; Mismatches 93; Indels 95; Gaps 14.

10 SPG-ATBEDYSVGSWYIDEFG-GXELDPEG-----VP-----40

Db 406 TPGDAITSLDFGSGNFDEPSAYGKALIPQGPVWVSGLRHSAQSESGSLPAFTVPOADPT 465

Qy 41 -----SCHTSTPRGLYHACLAS-LSIIYL-LLMLVLRRR----- 73

Db 466 GNTTPSASSSSSCGTPSSBEDYTSLENPLKIEYIEGLFMNDPEPRDPSGTOTDFSLD 555

Qy 74 -----QIMPDCVRGRPGILSPFVDPLACDRPRAVPAAVFVLLSLCLLPD 119

Db 526 ETLPAPIPMGDEDFQLSPIC-----PEEPLMPESPQPTQHCSTYTSIFQPLTPG 576

Qy 120 EDALPFLTLASAPSO--DGKTEAPRGAMKILGFHYALTYPLAACATAGHTAAHLG-- 175

Db 577 ATHGPFF-LDKYPPQLRESRTSESH--WPMSSIFPDAGSGSLSPCCGQASTPLSSWGR 633

Qy 176 STLSS-----AHLGVQVW-----QRAE 192

Db 634 SNTQMPDPDPLHFGPTKMPVYGDSAE 659

```
Search completed: March 13, 2003, 11:30:20
Job time : 10.4561 secs
```

QY 10 SPG-ATEDYSYGSWYIDEPQG-GXELOPEGE-----VP----- 40

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:27:17 ; Search time 11.391 Seconds

(Without alignments)
1679.148 Million cell updates/sec

Title: US-09-816-653A-2

Perfect score: 1061

Sequence: 1 MSQPAGNQTSPGATEDSYG.....MAHLGVQVWQRAECQVPR1 199

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86.5	8.2	757	2 F87304	beta-N-acetylhexos
2	85.5	8.1	380	2 S35785	glycoprotein I - b
3	84.5	8.0	335	2 T36304	probable anthranil
4	84	7.9	913	2 A48280	receptor tyrosine
5	83	7.8	1172	2 T36053	probable ABC-type
6	80.5	7.6	385	2 E83506	probable MFS trans
7	80	7.5	433	1 VHYUPH	nucleocapsid prote
8	80	7.5	876	2 A49508	protein-tyrosine k
9	80	7.5	921	2 G02326	transcription fact
10	79.5	7.5	249	1 S47153	type IV prelinin p
11	79.5	7.5	577	2 S77715	cytochrome c-type
12	79	7.4	463	2 C47301	virB6 homolog - Bo
13	79	7.4	910	2 A51137	tyrosine kinase re
14	78.5	7.4	477	2 H83588	probable MFS trans
15	78.5	7.4	532	2 JC5412	epidermal growth f
16	78.5	7.4	3519	2 S45048	polyketide synthas
17	78	7.4	419	2 JQ2254	latency1-diphospha
18	77.5	7.3	579	2 S35237	cytochrome c-type
19	77	7.3	521	2 AF3287	chiamine transport
20	76.5	7.2	321	2 B48266	protein-tyrosine k
21	76	7.2	394	2 A55045	probable 3-hydroxy
22	76	7.2	666	2 F83340	hypothetical prote
23	76	7.2	709	2 F75584	hypothetical prote
24	75.5	7.1	400	2 T35334	probable membrane
25	75.5	7.1	454	2 T10525	cyclin B1d-11 - ye
26	75.5	7.1	655	1 A42420	L-iduronidase (EC
27	75	7.1	211	2 T09498	cytochrome c bioge
28	75	7.1	230	2 S72714	Lept1170_F2_64 pro
29	75	7.1	668	2 T01685	crpl protein - mai

30	74.5	7.0	425	2 C87619	hypothetical prote
31	74	7.0	320	2 A82449	conserved hypothet
32	73.5	6.9	233	2 T47136	hypothetical prote
33	73.5	6.9	300	2 D83487	hypothetical prote
34	73.5	6.9	636	2 J00047	class I cytochrome
35	73.5	6.9	1174	2 A06663	probable pyruvate-
36	73.5	6.9	1201	2 A03007	hypothetical prote
37	73	6.9	411	2 D72588	hypothetical prote
38	73	6.9	473	2 T03611	cyclin, B-type - c
39	73	6.9	633	2 F84564	probable protein k
40	73	6.9	739	2 A83015	primosomal protein
41	73	6.9	786	2 F70735	hypothetical prote
42	72.5	6.8	176	2 B83199	hypothetical prote
43	72.5	6.8	529	2 S75950	hypothetical prote
44	72.5	6.8	258	2 C70545	hypothetical prote
45	72.5	6.8	636	1 VCVWFS	env polyprotein -

ALIGNMENTS

RESULT 1

F87304 beta-N-acetylhexosaminidase, probable [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #ext_change 20-Apr-2001

C:Accession: F87304

R:Nieman, W.C.; Feildblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.;

B.; Laub, M.T.; Debony, R.T.; Dodson, R.J.; Durkin, A.S.; Gwin, M.L.; Haft, D.H.; Kolton;

N.J.; Emolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MOID:21173698; PMID:11259647

A:Accession: F87304

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1757 <SNO>

A:Cross-references: GB:AE005673; NID:g13421618; PIDN:AAK2434.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0447

Query Match 82% Score 86.5; DB 2; Length 757;
Best Local Similarity 30.6%; Pred. No. 3.4;
Matches 49; Conservative 16; Mismatches 60; Indels 35; Gaps 9;

Qy	62	VLLLLMLVR-RQLWPDVGRGRLSPVDFLAGDRP-----RAVPA 103
Db	65	VALQLADLIQRGRFRKVGEPADAIY--LTRRGPAGEAYKLDINAKGATIAAKRA 122
Qy	104	AVFVLLSSICLLIPDEDAFLTLASPSODGKTEPRGAMKILGFHYAALYY----P 159
Db	123	GLFGAMSLMQLATPDEAKGFVALLAASIE----DAVRFMR--GLMVSARHYOSLDT 175
Qy	160	LAACATAGTAHLLGSLTSMHLLGVQVWQRAECQVPR1 199
Db	176	LKAVIDA--MAHKL-NTFHHLLVDDGW-RLRIKTKPKL 211

RESULT 2

S35785 glycoprotein I - bovine herpesvirus 1

C:Species: bovine herpesvirus 1

C:Date: 09-Jun-1994 #sequence_revision 12-May-1995 #ext_change 24-Nov-1999

C:Accession: S35785

R:Audonnet, J.

submitted to the EMBL Data Library, June 1993

A:Reference number: S35782

A:Accession: S35785

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-380 <AUD>

A:Cross-references: EMBL:Z23068; NID:g1312185; PIDN:CAA80605.1; PID:g1312189

C:Superfamily: pseudotabes virus glycoprotein gp63

C:Keywords: glycoprotein

Query Match 8.1%; Score 85.5; DB 2; Length 380;

Best Local Similarity 27.6%; Pred. No. 2.1;

Matches 37; Conservative 12; Mismatches 38; Indels 47; Gaps 6;

QY 4 PANGQTSFGATEDSYSGWYIDEPQGXELQPEGEVPSCHTSPGLYHACLASLSTLS 59

DB 256 PANGTASPAASPSRAFAA-----AAPAAAAQPAQDTPA-----RFRQLASTLVPLC 302

QY 60 ILVLLLLAM-----LVRRQL-----WPCVGRGPGCLPSV 90

DB 303 VLVLVLLALCAATVNCALRRRLPCRRRYVKPRTCAACSGTCACRPGCAAPAPATV 362

QY 91 DFLAGDRPRAVPAA 104

DB 363 VAL-GSRPKAPPLA 375

RESULT 3

probable anthranilate phosphoribotransferase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000

C:Accession: J36304

R:Saunders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, March 1999

A:Reference number: 221604

A:Accession: J36304

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-335 <SAU>

A:Cross-references: EMBL:AL035654; PIDN:CAB36583.1; GSPDB:GN00070; SCOEDB:SCB8.05C

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: trpD2; SCOEDB:SCB8.05C

C:Superfamily: anthranilate phosphoribosyltransferase; trpD homology

Query Match 8.0%; Score 84.5; DB 2; Length 335;

Best Local Similarity 30.8%; Pred. No. 2.2;

Matches 36; Conservative 10; Mismatches 42; Indels 29; Gaps 5;

QY 81 RGRPLPSVVDPLAGDRPRAVPAVFMVLSCLLPEDALPPLTLASAPSGKTEA 140

DB 247 RSRPG-----DIAGSD--RAVAAVLRVLAG-----GRGPARDVILLN 283

QY 141 PRGAWKITGLFHYALVYPLACATAGHTAHLIGSTLSWAHLGVQVWQRAECPQV 197

DB 284 AAAALRVAGVAGTWSDGRLAASAVDGAAGLLD--RMAHAS--WQADLVEVP 334

RESULT 4

A48280

receptor tyrosine kinase - human

C:Species: Homo sapiens (man)

C:Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 24-Sep-1999

C:Accession: A48280

R:Johnson, J.D.; Edman, J.C.; Rutter, W.J.

Proc. Natl. Acad. Sci. U.S.A. 90, 5677-5681, 1993

A:Title: A receptor tyrosine kinase found in breast carcinoma cells has an extracellular

A:Reference number: A48280; MUID:93296201; PMID:8390675

A:Accession: A48280

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-913 <RES>

A:Cross-references: GB:L11315; NID:G403386; PIDN:AAA02866.1; PID:G403387

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; discoidin I amino-ter

C:Keywords: ATP

F:J36135/Domains: discoidin I amino-terminal homology <DNI>

F:608-912/Domains: protein kinase homology <KIN>

F:616-624/Region: protein kinase ATP-binding motif

Query Match 7.9%; Score 84; DB 2; Length 913;

Best Local Similarity 22.7%; Pred. No. 7.1;

Matches 39; Conservative 20; Mismatches 73; Indels 40; Gaps 7;

QY 7 NOTSGATEDSYSGWYIDEPQ---GGXELQPEGEVPSCHT-SIPGLYHACLASLSTLS 61

DB 370 NNSSPALGTFPPAPWMPGPPPTWFSLELEPRQGVAAEGSPFALLIGCVAILLL 429

QY 62 VLLLLAMLV-----RRQLMPD-----CVGRPLSPVDFL----- 93

DB 430 LLLITALLMLRMLWRRLLSKAERRYLEELTVHLSVPGDTLLINRRPGRPREPPYOEBRP 489

QY 94 AGDRPRAVPAVFMVLSCLLPEDALPPLTLASAPSGDKTEAPRGAW 145

DB 490 RGNPPHSAFC-----VPNGSALLSNPAVRLLLATYARPPRGPC--PPTPM 534

RESULT 5

T36053

probable ABC-type transport protein cydC - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C:Accession: J36053

R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M

submitted to the EMBL Data Library, November 1998

A:Reference number: 221595

A:Accession: J36053

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1172 <SAU>

A:Cross-references: EMBL:AL03435; PIDN:CAA2219.1; GSPDB:GN00070; SCOEDB:SCD78.14

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: cydC; SCOEDB:SCD78.14

C:Superfamily: Mycobacterium tuberculosis probable ABC transporter Rv0194; ATP-binding c

Query Match 7.8%; Score 83; DB 2; Length 1172;

Best Local Similarity 31.5%; Pred. No. 12;

Matches 45; Conservative 12; Mismatches 68; Indels 18; Gaps 5;

QY 55 LASSIIVLLLLAMLVRRQLMPCVGRGCLPSVDFLAGDRPRAVPAVFMVLSLSC 114

DB 874 LAABEAVLGRLAVQYRQ--VRRSAERYVEVLDAPFVRREPEVRQAPATPFPVVEGLA 932

QY 115 LLLP--DEDALPPLTL-----ASAPSGDKTEAPRGAWKITGLFHYALVYPLA-- 161

DB 933 ARHNGQRDLALAGLDLLEGRRAVAVVPGSGSKTTL--AYVLLRFLDSAAAGSYTLAGV 989

QY 162 -ACATAGHTAHLIGSTLSWAHL 183

DB 990 DAYVALAGDVRRIVGLCAODAH 1012

RESULT 6

E83506

probable MFS transporter Pali08 [imported] - Pseudomonas aeruginosa (strain PAO1)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: E83506

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bro

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kass, A.; Lapid, K.; Lam,

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: E83506

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-385 <STO>

A:Cross-references: GB:AE004541; GB:AE004091; NID:G9947024; PIDN:AA04497.1; GSPDB:GN001

A:Experimental source: strain PAO1

C:Genetics:

A:Gene: Pali08

Query Match 7.6%; Score 80.5; DB 2; Length 385;
Best Local Similarity 29.0%; Pred. No. 6.1; Mismatches 57; Indels 61; Gaps 13;
Matches 54; Conservative 14; Mismatches 57; Indels 61; Gaps 13;
27 POGXELQEGEVPSCHTIPPGLYHACLASLSTLVLLLTAMVRRRLMPDQVGR-- 84
Db 148 PLGGOWLADJGHH-----AVFATLAVTLASLMAALV---PAMPET---RPLL 189
QY 85 -GLPSPVDF-----LAGDRP---RAVPAVPMVLLSLC-----LLLPDEDALP---LT 127
Db 190 AGTEPEPATLAFRRVLDRPLQTRALLVAVLVFSEFYAAGPVMVDLPGLGFWIGLA 249
QY 128 LASAPSDG--KTEAPRGAWKI-----LGLFYAALYPLAACATAGTAHLIGST--- 177
Db 250 IAIIGSGALLNRRRLR--TMSARVRVGL-----ALAAAGTQOTLLAAVGYA 297
QY 178 --LSMA 181
Db 298 EGLVYA 303

RESULT 7

VHVJPH nucleocapsid protein N - Prospect Hill virus (strain PHV-1)

N:Alternate names: nucleoprotein N

C:Species: Prospect Hill virus

C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 24-Jul-1997

C/Accession: A34681

R:Parrington, M.A.; Kang, C.Y.

Virology 175, 167-175, 1990

A:Title: Nucleotide sequence analysis of the S genomic segment of Prospect Hill virus: C

A:Reference number: A34681; MUID:90177212; PMID:2309440

A:Accession: A34681

A:Molecule type: genomic RNA

A:Residues: 1-433 <PAR>

C:Genetics:

A:Gene: N

A:Map position: segment S

C:Superfamily: Hantaan virus nucleocapsid protein

C:Keywords: nucleocapsid; nucleoprotein

Query Match 7.5%; Score 80; DB 1; Length 433;

Best Local Similarity 24.1%; Pred. No. 7.7;

Matches 35; Conservative 21; Mismatches 55; Indels 34; Gaps 6;

QY 23 YIDPEOGXELQEGEVPSCHTIPPGLYHACLASLSTLVLL---LAMVRRRLMPD 78
Db 178 YVSMPTAOSTMKAEBLTPGRFRITVGLPFAQIMANITSPWGVIGFAFFVKD---WAD 234
QY 79 CVRG-----RQLPS-PVDPLAGDRPRAVPAVPMVLLSLCL---LPDE 120
Db 235 KVAPLDQCKPFLKAPRPGQAGAEFLSSIR-----AYLMNRQAVDETHLPDI 285
QY 121 DALPFLTLASAPSDGKTEAPRGAW 145
Db 286 DALVELAAGDPTLPDLSLENPHAM 310

RESULT 8

A49508 protease-tyrosine kinase (EC 2.7.1.112) crk precursor - human

C:Species: Homo sapiens (man)

C>Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 04-Feb-2000

C/Accession: A49508; 138358; S37402

R:Di Marco, E.; Cutuli, N.; Guerra, L.; Cancedda, R.; De Luca, M.

J. Biol. Chem. 268, 24290-24295, 1993

A:Title: Molecular cloning of crk, a novel crk-related putative tyrosine kinase receptor

A:Reference number: A49508; MUID:94043265; PMID:8226977

A:Accession: A49508

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-876 <DIA>

A:Cross-references: EMBL:X74979; NID:G400462; PIDN:CAA52915.1; PID:G400463

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; discoidin I amino-term
C:Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase
F:30-185/Domain: discoidin I amino-terminal homology <DNI>
F:571-875/Domain: protein kinase homology <KIN>
F:579-587/Region: protein kinase ATP-binding motif

Query Match 7.5%; Score 80; DB 2; Length 876;
Best Local Similarity 20.6%; Pred. No. 16;
Matches 42; Conservative 26; Mismatches 74; Indels 62; Gaps 9;

QY 7 NOTSPATDYSYGWYIDPEO---GGXELQEGEVPSCHT-SIPPGLYHACLASLSTL 61
Db 370 NNSSPALGCTFPAPWPPGPPPTNPSLELPRQGPVAKAGSPTALIGCTVAIILL 429
QY 62 VLLILMLV-----RRQLMPD-----CVGRPGILSPVDPLAGDR 97
Db 430 LLLIILMLRLHWRRLSKAERRVLEELTVHLSPGDTILINRPPRPPPY-QEPR 408
QY 98 PRAVPAVPMVLLSLC-----LLPDEDALPFLTLASAPSDGKTEAPRGAWKI 147
Db 489 PRGNPPH-----SAPCVNGSAYSGDYMEPEKAPALL---PPPPN----- 527
QY 148 LGLFYAALYPLAACATAGTA 171
Db 528 -SVPHYADIVTLQGTGNTYA 550

RESULT 9

transcription factor NFAT1 isoform B - human

C:Species: Homo sapiens (man)

C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999

C/Accession: G02326

R:Juo, C.

submitted to the EMBL Data Library, December 1995

A:Reference number: H01056

A:Accession: G02326

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-921 <LUO>

A:Cross-references: EMBL:U43341; NID:G1353773; PIDN:AAC50886.1; PID:G1353774

C:Genetics:

A:Gene: NFAT1

Query Match 7.5%; Score 80; DB 2; Length 921;

Best Local Similarity 23.5%; Pred. No. 17;

Matches 57; Conservative 30; Mismatches 76; Indels 80; Gaps 15;

QY 3 QPAGNOSPQATDYSYGWYIDPEOGX-----ELOPEGEV-PSCHTIPPGLYHAC 54
Db 240 RPASRSSSPGAKRRHSCAEALVLPQASPORSRSPQSSHYAPQDHGS-PAG--YPP 296
QY 55 LASLSTLVLLTAMVRR-----RLW-----PDVGRGP---GLP---SPVDPLA--- 94
Db 297 VAGSAVMDALNSLATSPGIPPKMKWTSPPDPSPVSAAPSKAGLPRIYPAVDFLPCE 356
QY 95 -GDRPRAVPAVPMVLLSLCLLLPDEDALPFLTL---ASRP-----SQDG----- 136
Db 357 QGERNSAPESILVPPPTWPKPLVP--AIPICSIPTVASLPLEWPLSSGSGSYELRIE 413
QY 137 -----KTEAPRGAWKI-LGLFYAALYPLAACATAGTAHLIGSTLSMAHLGV 185
Db 414 VQPKRHRAHYETBSRGAVK-----APTGHVTVQLHG--YMNKPLGL 456
QY 186 QVW 188
Db 457 QIF 459

RESULT 10

S47153 type IV prepilin peptidase (EC 3.4.99.-) BfpP - Escherichia coli

N:Contains: type IV pilin N-methyltransferase (EC 2.1.1.-)

C:Species: Escherichia coli
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #ext_change 16-Jun-2000
 C:Accession: S47153; S70974
 R:Zhang, H.; Lory, S.; Domeneberg, M.S.
 Submitted to the EMBL Data Library, June 1994
 A:Description: A plasmid encoded prelipin peptidase from enteropathogenic Escherichia coli
 A:Reference number: S47153
 A:Accession: S47153
 A:Molecule type: DNA
 A:Residues: 1-249 <ZHA>
 A:Cross-references: EMBL:Z24464; NID:g499386; PIDN:CAA84229.1; PID:g499387
 R:Stone, K.D.; Zhang, H.Z.; Carlson, L.K.; Domeneberg, M.S.
 Mol. Microbiol. 20, 325-337, 1996
 A:Title: A cluster of fourteen genes from enteropathogenic Escherichia coli is sufficient
 A:Reference number: S70966; MUID:96310370; PMID:8733231
 A:Accession: S70974
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-249 <STO>
 A:Cross-references: EMBL:Z68186; NID:g112239; PIDN:CAA92334.1; PID:g1122408
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1995
 A:Genetics:
 A:Gene: dppp
 C:Superfamily: type IV prelipin peptidase
 C:Keywords: hydrolase; methyltransferase; S-adenosylmethionine

Query Match 7.5%; Score 79.5; DB 1; Length 249;
 Best Local Similarity 25.8%; Pred. No. 4.8;
 Matches 41; Conservative 11; Mismatches 62; Indels 45; Gaps 6;
 QY 22 WYIDPQGXELQPEGEVPSCHTSIP-----PGLYHACLASLS 59
 Db 60 WYLLLPVLQYFL-CRCECGYCHAKVPRVPLTEFLICGVCVLIIFVPLGRLVDIVLSL 118
 QY 60 ILVLLIAMLVRRRQWPCVGR-----RGLPSPVDFLAGDRPRAVPAVFMVLSL 111
 Db 119 FLCLVFLALIDLRNMLPCVCTVPLFWAGMITPGFASDDKIFG---AFTGLIMYISM 174
 QY 112 SLCLLPDEDALPFLTLGAPSDQKTEAPRGAMKILGL 150
 Db 175 KLVSAIRKEDVAGGDIALATA-----AGAW---LGI 203

RESULT 11

S77715
 cytochrome c-type biogenesis protein homolog - evening primrose mitochondrion
 C:Species: mitochondrion Oenothera villarica (evening primrose)
 C:Date: 06-Sep-1996 #sequence_revision 09-May-1997 #ext_change 24-Sep-1999
 C:Accession: S77715; S35238; S77730
 R:Schuster, W.; Combettes, B.; Flioger, K.; Brennicke, A.
 Mol. Gen. Genet. 239, 49-57, 1993
 A:Title: A plant mitochondrial gene encodes a protein involved in cytochrome c biogenesis
 A:Reference number: S35237; MUID:93288005; PMID:8369979
 A:Accession: S77715
 A:Molecule type: mRNA
 A:Residues: 1-577 <SCH>
 A:Cross-references: EMBL:X69555
 A:Note: the source is designated as Oenothera berteriana
 A:Note: 10-Ser, 13-Leu, 33-Leu, 46-Cys, 51-Ser, 86-Trp, 88-Leu, 106-Val, 112-Tyr
 -Cys, 259-Leu, 268-Leu, 306-Trp, 325-Phe, 368-Leu, 424-Trp, 433-Leu, 439-Tyr
 A:Note: this represents all RNA editing sites; individual clones show partial RNA editing
 A:Accession: S35238
 A:Molecule type: DNA
 A:Residues: 1-9, 'P', 11-12, 'P', 14-32, 'P', 34-45, 'S', 47, 'R', 49-50, 'P', 52-85, 'R', 87, 'P', 89-1
 'P', 237-238, 'S', 240-243, 'P', 245-246, 'R', 248-251, 'R', 253-258, 'S', 260-262, 'P', 264-267, 'S'
 2-466, 'R', 468-480, 'S', 482-487, 'L', 489-577 <SCF>
 A:Cross-references: EMBL:X69555
 R:Schuster, W.
 Submitted to the EMBL Data Library, February 1992
 A:Reference number: S77730
 A:Accession: S77730
 A:Molecule type: DNA
 A:Residues: 1-9, 'P', 11-12, 'P', 14-32, 'P', 34-45, 'S', 47, 'R', 49-50, 'P', 52-85, 'R', 87, 'P', 89-1

'P', 237-238, 'S', 240-243, 'P', 245-246, 'R', 248-251, 'R', 253-258, 'S', 260-262, 'P', 264-267, 'S'
 2-466, 'R', 468-480, 'S', 482-487, 'L', 489-577 <SCF>
 A:Cross-references: EMBL:X69555; NID:g14026; PIDN:CAA9287.1; PID:g14027
 C:Genetics:
 A:Genome: mitochondrion
 C:Superfamily: mitochondrial cycK-related protein
 C:Keywords: mitochondrion; RNA editing

Query Match 7.5%; Score 79.5; DB 2; Length 577;
 Best Local Similarity 22.9%; Pred. No. 12;
 Matches 25; Conservative 16; Mismatches 49; Indels 17; Gaps 2;
 QY 13 ATEDSYSGWYIDPQGXELQPEGEVPSCHTSIPGLYHACLASLSILVLLIAMLVRR 72
 Db 438 AVHELGMGGMFWBPVENASFMFWLATACIHVSILPLHSCITLLINIVFLC----- 490
 QY 73 RQLWPCVGR-----RGLPSPVDFLAGDRPRAVPAVFMVLSLCLL 117
 Db 491 -----CYLGFSTRSGLLASVHSFATDTRGIFLMRFLLMTGISML 533

RESULT 12

C47301
 virB6 homolog - Bordetella pertussis
 C:Species: Bordetella pertussis
 C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #ext_change 20-Sep-1999
 C:Accession: C47301
 R:Wetzel, A.A.; Johnson, F.D.; Burns, D.L.
 Proc. Natl. Acad. Sci. U.S.A. 90, 2970-2974, 1993
 A:Title: Molecular characterization of an operon required for pertussis toxin secretion.
 A:Reference number: A47301; MUID:93219406; PMID:8464913
 A:Contents: BP338
 A:Accession: C47301
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-463 <MB1>
 A:Note: sequence extracted from NCBI backbone (NCBIN:128775, NCBI:P:128778)
 C:Superfamily: Bordetella pertussis virB6 homolog

Query Match 7.4%; Score 79; DB 2; Length 463;
 Best Local Similarity 26.3%; Pred. No. 10;
 Matches 54; Conservative 18; Mismatches 71; Indels 62; Gaps 11;

QY 5 AGNQTSPGATEDSYSGWYIDPQGXELQPEGEVPSCHTS---IPGLYHACLASLS-- 59
 Db 220 AANDANGSAVMNMLCMYVASAGW-----LCALASLLIVPGLVTLTSLGLP 270
 QY 60 ILVLLIAMLVRRRQWPCV-----RGRPLSPVDFLAGDRPRAVPAV----- 105
 Db 271 FLVLLLPALQRTNMAVGLVAVFMALGTPAVGLSDVLAAGALPAGLPQRFATDPLR 330
 QY 106 FMVLLSLC---LLPDEDALPFLTLAS-----PSQDKTEAPRGAMKILG 149
 Db 331 STMIAATLCATATLML-----LTLVPLASSVAGLRRRLPMAHPGLAQAHQA----- 380
 QY 150 LFHYALVY---PLAACATAG-HTA 170
 Db 381 -----AAQVAPRPAAAAAAGPHQA 401

RESULT 13

A53137
 tyrosine kinase receptor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #ext_change 24-Sep-1999
 C:Accession: A53137
 R:Sanchez, M.P.; Tapley, P.; Saini, S.S.; He, B.; Pulido, D.; Barbacid, M.
 Proc. Natl. Acad. Sci. U.S.A. 91, 1819-1823, 1994
 A:Title: Multiple tyrosine protein kinases in rat hippocampal neurons: isolation of Pck-1
 A:Reference number: A53137; MUID:94173920; PMID:8127887
 A:Accession: A53137
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA

A:Residues: 1-910 <RES>
A:Cross-references: GB:L26525; NID:g432480; PID:g432481
C:Genetics:
A:Gene: Plk-3
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; discoidin I amino-terminal
C:Keywords: ATP
F:31-186/Domains: discoidin I amino-terminal homology <DN1>
F:605-909/Domains: protein kinase homology <KIN>
F:613-621/Region: protein kinase ATP-binding motif

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:27:42 ; Search time 10.2538 Seconds
(Without alignments)
571.023 Million cell updates/sec

Title: US-09-816-653A-2

Perfect score: 1061

Sequence: 1 MSQPNAGNQTSPGATEDYSYG.....MAHIGVQWRACPCQVPKI 199

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*

2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*

3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*

4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*

5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*

6: /cgn2_6/ptodata/2/1aa/backfltest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85.5	8.1	380	4 US-08-924-345-2	Sequence 2, Appli
2	84	7.9	919	1 US-08-336-343A-2	Sequence 2, Appli
3	83	7.8	913	1 US-08-445-640-4	Sequence 4, Appli
4	83	7.8	913	3 US-08-170-558-4	Sequence 4, Appli
5	83	7.8	913	3 US-08-447-314-4	Sequence 4, Appli
6	83	7.8	913	3 US-08-445-461-4	Sequence 4, Appli
7	80	7.5	433	3 US-09-106-075A-86	Sequence 86, Appli
8	80	7.5	699	5 PCT-US94-07297-39	Sequence 39, Appli
9	80	7.5	921	1 US-08-396-479B-2	Sequence 2, Appli
10	80	7.5	921	1 US-08-818-823-2	Sequence 2, Appli
11	79	7.4	505	1 US-08-221-750A-5	Sequence 5, Appli
12	78.5	7.4	3519	4 US-09-428-517-4	Sequence 4, Appli
13	76.5	7.2	875	1 US-08-785-241-5	Sequence 5, Appli
14	75	7.1	331	2 US-08-986-217-7	Sequence 7, Appli
15	73.5	6.9	378	4 US-09-482-180A-2	Sequence 2, Appli
16	73.5	6.9	578	1 US-08-653-740-3	Sequence 3, Appli
17	73.5	6.9	578	2 US-09-073-594-3	Sequence 3, Appli
18	73.5	6.9	578	1 US-09-275-925-3	Sequence 3, Appli
19	73.5	6.9	636	1 US-08-653-740-5	Sequence 5, Appli
20	73.5	6.9	636	2 US-09-073-594-5	Sequence 5, Appli
21	73.5	6.9	636	3 US-09-275-925-5	Sequence 5, Appli
22	72	6.8	652	4 US-09-310-463-4	Sequence 4, Appli
23	72	6.8	652	4 US-08-842-248A-4	Sequence 4, Appli
24	72	6.8	694	4 US-09-440-325A-1	Sequence 4, Appli
25	69.5	6.6	369	4 US-08-463-509B-4	Sequence 4, Appli
26	69.5	6.6	369	5 PCT-US95-05616-4	Sequence 4, Appli
27	69.5	6.6	372	4 US-08-462-509B-6	Sequence 6, Appli

28	69.5	6.6	377	5 PCT-US95-05616-6	Sequence 6, Appli
29	69.5	6.6	389	2 US-08-846-705-2	Sequence 2, Appli
30	69.5	6.6	402	4 US-08-462-509B-2	Sequence 2, Appli
31	69.5	6.6	402	5 PCT-US95-05616-2	Sequence 2, Appli
32	69.5	6.6	425	3 US-08-846-704-2	Sequence 2, Appli
33	69.5	6.6	650	4 US-09-310-463-2	Sequence 2, Appli
34	69.5	6.6	650	4 US-08-842-248A-2	Sequence 2, Appli
35	69.5	6.6	1205	1 US-07-908-245-2	Sequence 2, Appli
36	68.5	6.5	402	3 US-08-846-704-4	Sequence 4, Appli
37	68.5	6.5	425	4 US-09-479-128-2	Sequence 4, Appli
38	68	6.4	591	1 US-08-188-582-7	Sequence 7, Appli
39	68	6.4	591	1 US-08-646-715-7	Sequence 7, Appli
40	68	6.4	1064	3 US-08-726-214-8	Sequence 8, Appli
41	67.5	6.4	373	4 US-09-254-077A-12	Sequence 12, Appli
42	67.5	6.4	2864	4 US-08-469-260A-394	Sequence 394, Appli
43	67	6.3	389	5 PCT-US93-11703-24	Sequence 24, Appli
44	67	6.3	389	2 US-08-789-354-2	Sequence 2, Appli
45	67	6.3	389	3 US-09-110-937-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-924-345-2
Sequence 2, Application US/08924345
Patent No. 6224878
GENERAL INFORMATION:
APPLICANT: LEUNG-TACK Patricia
APPLICANT: LEGASTROIS Isabelle, Christine, Marie-Andree
APPLICANT: AUDONNET Jean-Christophe, Francis
APPLICANT: RIVIERE Michel, Emile, Albert
TITLE OF INVENTION: Mutants and vaccines of the infectious
TITLE OF INVENTION: Bovine Rhinotracheitis virus
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: LARSON AND TAYLOR
STREET: 727 SOUTH TWENTY-THIRD STREET
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/924,345
FILING DATE: 04-SEP-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/199,172
FILING DATE: 09-AUG-1994
APPLICATION NUMBER: FR 92 07930
FILING DATE: 26-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: SARRO, THOMAS P.
REGISTRATION NUMBER: 19396
REFERENCE/DOCKET NUMBER: XI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 920-7200
TELEFAX: (703) 892-8428
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 380 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-924-345-2

Query March 8.1%; Score 85.5; DB 4; Length 380;
Best Local Similarity 27.6%; Pred. No. 0.18;

Matches 37; Conservative 12; Mismatches 38; Indels 47; Gaps 6;

QY 4 PAGNQTSPGATEDSYSGWYIDEPQ---GGXELQPEGEVPSCHT-SIPPGLYHACLASISIL 59
DB 256 PAGNQTSPGATEDSYSGWYIDEPQ---GGXELQPEGEVPSCHT-SIPPGLYHACLASISIL 302
QY 60 IIVLLILLAM-----LVRRQL-----WPCVGRPGLSVDF 90
DB 303 IIVLLILLAM-----LVRRQL-----WPCVGRPGLSVDF 362
QY 91 DVLADGPRAPVAPAA 104
DB 363 VAL-GSRPRAPPLA 375

RESULT 2

US-08-336-343A-2
Sequence 2, Application US/08336343A
Patent No. 5677144
GENERAL INFORMATION:
APPLICANT: Ullrich, Axel
TITLE OF INVENTION: CCK-2, A No. 5677144e1 Receptor Tyrosine Kinase
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,343A
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7683-065
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 919 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-336-343A-2

Query Match Best Local Similarity 7.9%; Score 84; DB 1; Length 919;

Matches 39; Conservative 20; Mismatches 73; Indels 40; Gaps 7;

QY 7 NOTSPGATEDSYSGWYIDEPQ---GGXELQPEGEVPSCHT-SIPPGLYHACLASISIL 61
DB 370 NNSSPALGCTFPAPWMPGPPPTNFSSLEPRGQGPVAKBESPTAILIGCVAILLL 429
QY 62 VLLILLAMLV-----RRQLMPD-----CVRGRPGLSVDF 93
DB 430 LLLIILMLMRLHWRRLSKAERVLLEELTVHLSVPGDTLLINRRGPPRPPEPPYQEBRP 489
QY 94 AGDRPRAVPAVFWVLSSCLLPDEDALPFLTLASAPSDGKTEAPRGAM 145
DB 490 RGNPPHSAPC---VPNGSALLSNPAIRLLATYARPPRGG---PPTPAW 534

RESULT 3

US-08-445-640-4
Sequence 4, Application US/08445640
Patent No. 5709858
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Baron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/445,640
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170558
FILING DATE: 20-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hassek, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C2
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-445-640-4

Query Match Best Local Similarity 7.8%; Score 83; DB 1; Length 913;

Matches 39; Conservative 20; Mismatches 73; Indels 40; Gaps 7;

QY 7 NOTSPGATEDSYSGWYIDEPQ---GGXELQPEGEVPSCHT-SIPPGLYHACLASISIL 61
DB 370 NNSSPALGCTFPAPWMPGPPPTNFSSLEPRGQGPVAKBESPTAILIGCVAILLL 429
QY 62 VLLILLAMLV-----RRQLMPD-----CVRGRPGLSVDF 93
DB 430 LLLIILMLMRLHWRRLSKAERVLLEELTVHLSVPGDTLLINRRGPPRPPEPPYQEBRP 489
QY 94 AGDRPRAVPAVFWVLSSCLLPDEDALPFLTLASAPSDGKTEAPRGAM 145
DB 490 RGNPPHSAPC---VPNGSALLSNPAIRLLATYARPPRGG---PPTPAW 534

RESULT 4

US-08-170-558-4
Sequence 4, Application US/08170558
Patent No. 6001621
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.

APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Baron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/170,558
FILING DATE: 20-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Haasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
TOPOLOGY: linear

US-08-170-558-4

Query Match 7.8%; Score 83; DB 3; Length 913;
Best Local Similarity 22.7%; Pred. No. 1.2; Mismatches 73; Indels 40; Gaps 7;
Matches 39; Conservative 20;

QY 7 NOTSPGATEDYSGSYIDEPQ---CGXELQEGEVPSCH-TSIPGLYHACLASLIL 61
DB 370 NNSSPALGCTFPAPWPPGPPPTNFSSLELPRGQGVAKBPGSTALILICLVAILL 429
QY 62 VLLILMLV-----RRQLMPD-----CVRGRPLSPVDL----- 93
DB 430 LLLIILMLRMLHWRLLSKAERVLSEELTVHLSVPGDTILLNRRGPREPPYQERPR 489
QY 94 AGDRPRAVPAVFMVLLSICLLIPDEDALPFLTLASAPSDGKTAPRGAW 145
DB 490 RGNPPHSAPC---VPGSALLLSNPAYRLLATYARPPRPG---PPTPAW 534

RESULT 5
US-08-447-314-4
Sequence 4, Application US/08447314
GENERAL INFORMATION:
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Baron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,314
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170558
FILING DATE: 20-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Haasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C1D2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
TOPOLOGY: linear

US-08-447-314-4

Query Match 7.8%; Score 83; DB 3; Length 913;
Best Local Similarity 22.7%; Pred. No. 1.2; Mismatches 73; Indels 40; Gaps 7;
Matches 39; Conservative 20;

QY 7 NOTSPGATEDYSGSYIDEPQ---CGXELQEGEVPSCH-TSIPGLYHACLASLIL 61
DB 370 NNSSPALGCTFPAPWPPGPPPTNFSSLELPRGQGVAKBPGSTALILICLVAILL 429
QY 62 VLLILMLV-----RRQLMPD-----CVRGRPLSPVDL----- 93
DB 430 LLLIILMLRMLHWRLLSKAERVLSEELTVHLSVPGDTILLNRRGPREPPYQERPR 489
QY 94 AGDRPRAVPAVFMVLLSICLLIPDEDALPFLTLASAPSDGKTAPRGAW 145
DB 490 RGNPPHSAPC---VPGSALLLSNPAYRLLATYARPPRPG---PPTPAW 534

RESULT 6
US-08-445-461-4
Sequence 4, Application US/08445461
Patent No. 6096527
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Baron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/445,461

FILED DATE: 22-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170558
FILING DATE: 20-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Haasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-445-461-4

Query Match 7.8%; Score 83; DB 3; Length 913;
Best Local Similarity 22.7%; Pred. No. 1.2;
Matches 39; Conservative 20; Mismatches 73; Indels 40; Gaps 7;

QY 7 NOTSPGATEDYSGSWYIDEPQ---GGXELQEGEVPSCH--TSIPGLYHACLASLTL 61
DB 370 NNSPFGALGTFPPAPMPWPPEPTNSSLELFRGQPVAKPEGSTALIGLVAILL 429
QY 62 VLLLLMLV-----RRRQLMPD-----CVRGRPGLPSPVDPL---- 93
DB 430 LLLIILMLRLMRLLSKAERRVLEELTVLHVSFGDTIILNNRPGREPPQGERPR 489
QY 94 AGDRPAVPAVFMVLLSLCLLPEDALPFLTLASASQSGKTEAPRGAW 145
DB 490 RGNPPHSAPC---VNGSALLSNPAVYLLATYARPPRGG---PPTPAW 534

RESULT 7
US-09-106-075A-86
Sequence 86, Application US/09106075A
Patent No. 6316250
GENERAL INFORMATION:
APPLICANT: Hjelle MD, Brian
TITLE OF INVENTION: Molecular Clones Producing Recombinant DNA Antigens of
FILE REFERENCE: 10312-801, Hjelle et al. (210312.0009)
CURRENT FILING DATE: US/09/106,075A
PRIOR APPLICATION NUMBER: 08/210,762
PRIOR FILING DATE: 1994-03-22
PRIOR APPLICATION NUMBER: 08/141,035
PRIOR FILING DATE: 1993-10-26
PRIOR APPLICATION NUMBER: 08/120,096
PRIOR FILING DATE: 1993-09-13
PRIOR APPLICATION NUMBER: 08/111,519
PRIOR FILING DATE: 1993-08-25
NUMBER OF SEQ ID NOS: 90
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 86
LENGTH: 433
TYPE: PRT
ORGANISM: Prospect Hill virus
US-09-106-075A-86

Query Match 7.5%; Score 80; DB 4; Length 433;
Best Local Similarity 24.1%; Pred. No. 0.91;
Matches 35; Conservative 21; Mismatches 55; Indels 34; Gaps 6;

QY 23 YIDFQGXELQEGEVPSCHTSIPGLYHACLASLTLVLL---LAMLVRRRLQMPD 78

DB 178 YVSMPLAOSTMAEELTYPERFTIVCGLPFAQIMARNIISPVMGVIGFAFFVKD---MAD 234
QY 79 CVRG-----PRGLPS-PVDFLAGDRPAPVPAVFMVLLSLCLL---LPDE 120
DB 235 KYKAFADQKCPFLKAEPPRGQPAGEAEFLSSIR-----ATLMNQAVLDETHLPDI 285
QY 121 DALPFLTLASAPSQSGKTEAPRGAW 145
DB 286 DALVEILASGDPTLPDPSLENPHAAW 310

RESULT 8
PCT-US94-07297-39
Sequence 39, Application PC/TUS9407297
GENERAL INFORMATION:
APPLICANT: Arai, Naoko
APPLICANT: Masuda, Esteban S.
APPLICANT: Tokumitsu, Hiroshi
TITLE OF INVENTION: PURIFIED COMPONENTS OF MAMMALIAN
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: John H. C. Blasdale, Schering-Plough Corporation, M-3-W
STREET: One Giralda Farms
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 07940-1000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh IIcx
OPERATING SYSTEM: System Software 7.1
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/07297
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/222,626
FILING DATE: 04-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,061
FILING DATE: 05-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/113,971
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/099,998
FILING DATE: 30-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/088,483
FILING DATE: 06-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Blasdale, John H. C.
REGISTRATION NUMBER: 31,895
REFERENCE/DOCKET NUMBER: DX0392K4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-822-7398
TELEFAX: 201-822-7039
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 699 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
PCT-US94-07297-39

Query Match 7.5%; Score 80; DB 5; Length 699;
Best Local Similarity 23.5%; Pred. No. 1.7;
Matches 57; Conservative 30; Mismatches 76; Indels 80; Gaps 15;

QY 3 QPAGNOTSPGATEDYSGSWYIDEPQGX-----ELOPEGEV-PSCHTSIPGLYHAC 54

```
Db 21 PASSSSSSGAKRRHSCAALVALPGASPORSRSPQSSSHVAPQDHGS--PAG--YPP 77
Qy 55 LASLSILVLLLLAMLVRR-----RQLW-----PDCVGRGP---GLP---SPVDFLA--- 94
Db 78 VAGSAVIMDALNSLATDPSGCIIPKMWKTSPPSPVSAAPSAKAGLPRIHYPAVEFLGPCE 137
Qy 95 -GDRRAVPAVAVFVWLSSCLLLPDEDALPFLT---ASAP-----SODG----- 136
Db 138 QGERNSAPESILLVPTWPKPLVP---AIPICSIPTVTALEPPLWPLSSGSGSYELRIE 194
Qy 137 -----KTEAPRGAMKILGLFHYALYYPLAACATGHTAAHLGSLTSMANHLGV 185
Db 195 VOPKPHRAHYETEGSRGAVK-----APTGHPPVQVLHG--YMNKPKLGL 237
Qy 186 QVW 188
Db 238 QIF 240

RESULT 9
US-08-396-479B-2
; Sequence 2, Application US/08396479B
; Patent No. 5612455
; GENERAL INFORMATION:
; APPLICANT: HOEY, Timothy
; TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396,479B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59450-1/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8771
; TELEFAX: (415) 494-8771
; TELEX: 210 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-396-479B-2

Query Match 7.5%; Score 80; DB 1; Length 921;
Best Local Similarity 23.5%; Pred. No. 2.5;
Matches 57; Conservative 30; Mismatches 76; Indels 80; Gaps 15;
```

```
Db 357 QGERNSAPESILLVPTWPKPLVP---AIPICSIPTVTALEPPLWPLSSGSGSYELRIE 413
Qy 137 -----KTEAPRGAMKILGLFHYALYYPLAACATGHTAAHLGSLTSMANHLGV 185
Db 414 VOPKPHRAHYETEGSRGAVK-----APTGHPPVQVLHG--YMNKPKLGL 456
Qy 186 QVW 188
Db 457 QIF 459

RESULT 10
US-08-818-823-2
; Sequence 2, Application US/08818823
; Patent No. 5708158
; GENERAL INFORMATION:
; APPLICANT: HOEY, Timothy
; TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,823
; FILING DATE: 14-MAR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/396,479
; FILING DATE: 02-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59450-1/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8770
; TELEFAX: (415) 494-8771
; TELEX: 210 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-818-823-2

Query Match 7.5%; Score 80; DB 1; Length 921;
Best Local Similarity 23.5%; Pred. No. 2.5;
Matches 57; Conservative 30; Mismatches 76; Indels 80; Gaps 15;
```

Db 414 VQPKHHRHAYETEGSRGAVK-----APTGGHPVVLHG-YMENKPIGL 456
QY 186 QVM 188
Db 457 QTF 459

RESULT 11

US-08-221-750A-5
Sequence 5, Application US/08221750A
Patent No. 5643747
GENERAL INFORMATION:
APPLICANT: Baker, Steven M.
APPLICANT: Deich, Robert A.
TITLE OF INVENTION: Genes for the Export of Pertussis
TITLE OF INVENTION: Holo toxin
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/221,750A
FILING DATE: 31-MAR-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/031,619
FILING DATE: 15-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: ACC93-01A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-221-750A-5

Query Match 7.4%; Score 79; DB 1; Length 505;
Best Local Similarity 26.3%; Pred. No. 1.5;
Matches 54; Conservative 18; Mismatches 71; Indels 62; Gaps 11;

QY 5 AGQTGPGATEDSYGSWYDEBQGXELQPEGEVPSCHTS---IPGLYHACIASIS-- 59
Db 220 AANDAFGSAVNWDLCLMIVAASNGW-----LCLASLLYVGLVTLISGL 270
QY 60 ILVLLLLAMLVRRROLMPDCV-----RGRQLPSVPVFLAGDRRAVPAV----- 105
Db 271 FLVLLLPALQRTNMLGLVALVPMALGTAVGLLSVLAGALPAGLPQFPATEPLR 330
QY 106 FMYVLSLC---LLPDEBALPFLITASA-----PSQDKTEAPGAKWILG 149
Db 331 STMLAATLCATATLMT-----LTLVPLASSVNGLRRLPMAHPGLAQAHROA----- 380
QY 150 LFHYAALY---PLACATAG-HTA 170
Db 381 ----AAQYARRPAAAAAAPHQA 401

RESULT 12

US-09-428-517-4
Sequence 4, Application US/09428517
Patent No. 6251636
GENERAL INFORMATION:
APPLICANT: Belkach, Mary C.
APPLICANT: Shan, Sanjay Krishnakant
APPLICANT: McDaniel, Robert
APPLICANT: Tang, Li
TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 30062-20029.00
CURRENT APPLICATION NUMBER: US/09/428,517
EARLIER FILING DATE: 1999-10-28
EARLIER APPLICATION NUMBER: 60/120,254
EARLIER FILING DATE: 1999-02-16
EARLIER APPLICATION NUMBER: 60/106,100
EARLIER FILING DATE: 1998-10-29
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 4
LENGTH: 3519
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Recombinant
OTHER INFORMATION: Oleandolide PKS
US-09-428-517-4

Query Match 7.4%; Score 78.5; DB 4; Length 3519;
Best Local Similarity 23.5%; Pred. No. 23;
Matches 54; Conservative 25; Mismatches 70; Indels 81; Gaps 12;

QY 11 PGATEDSYGSWYDEBQGXELQPEGEVPSCHTSIPGLYHACIASISLVLILLMLV 70
Db 2642 PALDEVLGGGLFVVPRG-----LADGGVAVRAVVAATVA 2676
QY 71 R-----RQIMPCVGRGRGLPSVPVFLAGDRR-----AVPA--AVFWL 109
Db 2677 RGEVSVVELDTPRDRRAVAAGR-GVSGVSFLSDDRHSEHPVPGLAASLVL 2735
QY 110 LSLC-----LLPDEBALPFLITASAPSDGKTEAPRGA--W---KILGLFH-- 152
Db 2736 AOLVLDLGRVSGPRMLVLTDA-----VAGPSDAGAVIDVQVQWFCGVLGIEHPE 2790
QY 153 --YAALYPLAA---CATAGTAAHLGSLSWAH-----GVQVWOR 190
Db 2791 LWGLJLDLPEVAPPEGSTCDHTYADLTAVASAGFEDQVAVRGSGWVR 2840

RESULT 13
US-08-785-241-5
Sequence 5, Application US/08785241
Patent No. 5695963
GENERAL INFORMATION:
APPLICANT: McKnight, Steven L.
APPLICANT: Russell, David W.
APPLICANT: Tian, Hui
TITLE OF INVENTION: Endothelial PAS Domain Protein
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/785,241
FILING DATE: 17-JAN-1997

```

: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: UTSD:1229
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 343-4341
: TELEFAX: (415) 343-4342
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 875 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-785-241-5

```

Query Match	7.2%;	Score 76.5;	DB 1;	length 875;
Best Local Similarity	22.9%;	Pred. No. 5.9;		
Matches 61;	Conservative 17;	Mismatches 93;	Indels 95;	Gaps 14;

[illegible]

Db	407	TRGDALISLDGSGNFDEPSAVGKALILPPGQWWSGLBSHSAQSESGSLPATTVEQADTP	466
Oy	41	-----SCHTSIPPGVTHACLAS-LSLTVL-LLLAMVRRR-----	73

Db 467 GNTTPSASSSSSCSPSPSPEDYSSLENDPKIEVLEKLFAMOTERDRPQSTQTDFSELDL 526
QY 74 -----QIWPDCYRGRGELSPVDVFLAGDRBRAVPAAVFWLLSSLCILLPD 119

Db 527 ETLAPYIMDGEDFGLSPIC-----PBEPLMBESQPPHOCHFMTMSI FQPLNPG 577

OY 120 EDALPELTLASAPG--DGKTEAPRGANKILCFHYALYLYPLAACATRGHNAALLG--- 175

120 EDAAPETLILASAPSV - - - GGTLEAPGANKLILCFLEFMAALYYFLAACATAGHFAAHLG -- 175
 176 STLSW-----AHLGVGVV-----ORAE 192
 Db 578 ATHGCFE-LDKXPKQLERKTESEH--WPMMSIFFDAGSKGSLPCCGCASTPLSSMGR 634

QY 176 STLSW-----AHLGVQVW----QRRE 192
| | | | | | | | | |
Db 635 SNTQWPPDPPLHFGPTKWPVGDQSAE 660
| | | | | | | | | |

RESULT 14
US-08-986-217-7
Sequence 7, Application US/08986217

```

; Sequence 7, Application US/08386217
; Patent No. 5914249
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti

```

APPLICANT: Lai, Preeti
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: CELL-CYCLE PHOSPHOPROTEINS
NUMBER OF SEQUENCES: 7

NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: San Diego
STATE: CA
ZIP: 92161

STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

```

? ZIP: 94304
?
? COMPUTER READABLE FORM:
?
? MEDIUM TYPE: Diskette
?
? COMPUTER: IBM Compatible
?
?

```

```

? COMPUTER: IBM compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: Fastseq for Windows Version 2.0
? CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/986.217

```

; APPLICATION NUMBER: US/08/986,217
 ;
 ; FILING DATE: Filed Herewith
 ;
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:

APPLICATION NUMBER :
FILING DATE :
ATTORNEY/AGENT INFORMATION :
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749

```

1 REFERENCE/DOCKET NUMBER:  PR-0452 US
2
3 TELECOMMUNICATION INFORMATION:
4
5 TELEPHONE: 650-855-0555
6
7 TELEFAX: 650-845-4166
8
9 INFORMATION FOR SEQ ID NO: 7:
10
11 SEQUENCE CHARACTERISTICS:
12
13     LENGTH: 331 amino acids
14     TYPE: amino acid
15     STRANDEDNESS: single
16     TOPOLOGY: linear
17
18 IMMEDIATE SOURCE:
19     LIBRARY: Genbank
20     CLONE: 2072296
21
22 US-08-986-217-7

```

Query Match	7.1%;	Score 75;	DB 2;	Length 331;
Best Local Similarity	22.7%;	Pred. No. 2.3;		
Matches 49;	Conservative 25;	Mismatches 84;	Indels 58;	Gaps 9;

QY	11	PGATDYSYGSWYIDEPGGKXELQPE---	GEVPSCHTISIPROLYNACLASLSILVLLLLA	67
23	PMGAEPMALGPTSPKPPAGAOLFPGFLIGDIPRTVTPQP	-----RPSLGIMEY-----	71	
		Matches 49; Conservative 25; Mismatches 84; Indels 58; Gaps 9;		

Db 23 PMGAEPMALSGSTPSKPKPAGAGQFLPGLGDIPIPVTPQP-----RPSLGIMEV---- 71

QY 68 MLVRRQLMPDCVGRGPGLPSP-----PV---DFLA---GDRPRAV-PAAVFWLL 110

```

Db      72  ----RSPHSGSGPPQVPLPTHKDKSGAPVRSYDVVASPOLGSTPENTRMRGASFSVLH 127
Qy     111 SSLCLLPDEDAI-----PFLTLASAPSDGKTEAPRGANKIIGL 150
      :      :      :      :      :      :      :      :      :      :

```

Db	128	TPLSGAIPSSPAVSPATIGSRKTTLS	PAQMDPPTYGDAITSDQDLD	--DTWVTEFG	184
Qy	151	FHYAALYPTLAACATAGCTAAHLIGSTISMAHLGQ			186

QY 151 FHVALYVPLACATVAGHTAAHLGGSTLSMAHLGQ 186
 Db 185 FQASASYILQFAQYGNIIKHWMSNNNGMNHFIQYQ 220

RESULT 15
US-09-482-180A-2
Sequence 2, Application US/09482180A
Description: 23 pages

```

:
:
: Sequence 2, Application US/09482180A
: Patent No. 6361985
:
: GENERAL INFORMATION:
:
: APPLICANT: Conklin, Darrell C.
: APPLICANT: Yamamoto, Gayle

```

APPLICANT: Yamamoto, Gayle
APPLICANT: Gao, Zeren
APPLICANT: Whitmore, Theodore E.
APPLICANT: Jaspers, Stephen

: APPLICANT: Jaspers, Stephen
 : TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING
 : TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOG, ZNSB6
 : FILE REFERENCE: 98-80
 : CURRENT APPLICATION NUMBER: US/09/482,180A

CURRENT APPLICATION NUMBER: US/09/482,180A
 CURRENT FILING DATE: 2000-01-12
 PRIOR APPLICATION NUMBER: 60/115,721
 PRIOR FILING DATE: 1999-01-12

```

;
; PRIOR FILING DATE: 1999-01-12
;
; NUMBER OF SEQ ID NOS: 20
;
; SOFTWARE: FastSeq for Windows Version 3.0
;
; SEQ ID NO 2
;
; LENGTH: 378

```

```

; LENGTH: 378
;
; TYPE: PRT
;
; ORGANISM: Homo sapiens
US-09-482-180A-2

```

Query Match	6.9%	Score 73.5	DB 4	Length 378
Best Local Similarity	25.0%	Pred. No. 4.1		
Matches 53: Conservative	16	Mismatches	82	Indels 61
		Gaps	11	

Matches 53; Conservative 16; Mismatches 82; Indels 61; Gaps 11;

Db 15 GRSGLTPG-----PAMLCRLCWLVSYSLVNLLGCLFLRKAKP--AGDPYAH 62

QY 88 SPVDPLAGDRR-----AVPAVPMVLISLSC---LLLP---DEDALP 124

D6 63 QP - FMAPPTDRHSRCPNNHTVSASISLSPRRHRLFLTYRHCRNFSILPEPGSGCKDTFL 120

Qy 125 FLTLASAPSDQKTEAPRGAWKILGLF---HYAALYPLACATAGHTTAHLG----- 175
 Db 121 LLAIKOPGHVERRAIRSTWGRVSGMARROKLVFLGVAGSH--PPAQLIAYESREF 178
 Qy 176 -STLSW-----AHLGVQWQRAECPQ 195
 Db 179 DDILQWDFTEDPFNLTKELHLQKRVVAACPQ 210

Search completed: March 13, 2003, 11:31:33
 Job time : 14.2538 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:25:47 ; Search time 25.0649 Seconds
(without alignments)
1057.929 Million cell updates/sec

Title: US-09-816-653A-2
Perfect score: 1061
Sequence: 1 MSQPAGNQTSFGATEDSYG.....MAHLGVVQMRACPEQVPXK 199

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

1: A_Geneseq_101002.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1048	98.8	487	23	AAO17374
2	1048	98.8	667	22	AAU04557
3	1048	98.8	667	22	AAU04557
4	1048	98.8	667	23	AAU78575
5	1048	98.8	667	23	AAU83631
6	1047	98.7	667	23	ABO5048
7	1043	98.3	662	23	ABO5047
8	1020	96.1	689	23	AAU91278
9	986.5	93.0	658	22	AAU04558
10	986.5	93.0	658	23	AAU78576

11	985.5	92.9	653	23	ABO5046
12	898	84.6	560	22	AAU91277
13	488	46.0	660	22	AAU91277
14	87.5	8.2	1070	22	AAU91277
15	85.5	8.1	380	15	ABG02155
16	84	7.9	563	18	AAW34674
17	84	7.9	624	21	AAW34674
18	84	7.9	624	21	AAW34674
19	84	7.9	624	21	AAW34674
20	84	7.9	624	21	AAW34674
21	84	7.9	624	21	AAW34674
22	80.5	7.6	130	20	AAW34674
23	80	7.5	433	18	AAW34674
24	80	7.5	876	23	AAW34674
25	80	7.5	876	23	AAW34674
26	80	7.5	882	18	AAW34674
27	80	7.5	921	17	AAW02248
28	80	7.5	1171	19	AAW85037
29	80	7.5	1181	19	AAW85038
30	79.5	7.5	485	22	ABG19631
31	79	7.4	505	18	AAW23275
32	79	7.4	699	16	AAW66875
33	78.5	7.4	3519	21	AAW92709
34	78	7.4	263	22	AAU50117
35	78	7.4	474	22	ABG09574
36	78	7.4	501	22	ABG13111
37	77.5	7.3	914	16	AAW71100
38	77	7.3	535	22	ABG03456
39	76.5	7.2	215	22	ABG15237
40	76.5	7.2	337	22	ABG15237
41	76.5	7.2	875	19	AAW37098
42	76	7.2	561	22	AAU45301
43	76	7.2	2189	22	AAW79017
44	75.5	7.1	466	23	AAW48999
45	75.5	7.1	740	22	ABG03831

ALIGNMENTS

RESULT 1
AAO17374 standard; Protein; 487 AA.
ID AAO17374
AC AAO17374;
XX 08-JUL-2002 (first entry)
XX 08-JUL-2002 (first entry)
XX Human retinoic acid responsive protein 53-57.
XX Human retinoic acid responsive protein 53-57; arrhythmia;
XX bronchial asthma; peptic ulcer; diabetes; cancer; cytoskeletal;
XX antiasthmatic; antidiabetic; endocrine; cardiac;
XX embryonic development disorder; gene therapy.
XX Homo sapiens.
XX WO200226808-A1.
XX 04-APR-2002.
XX 20-AUG-2001; 2001MO-CN01252.
XX 23-AUG-2000; 2000CN-0119715.
XX (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
XX Mao Y, Xie Y;
XX WPI; 2002-330093/36.
XX N-PSDB; AAL45968.
XX Human retinoic acid responsive protein 53.57 and encoding

FT polynucleotide, used in diagnosis and treatment of diseases e.g.,
 PT arrhythmia, asthma, diabetes and cancer -
 PS Claim 1, Page 32-33; 37pp; Chinese.
 XX
 CC The present invention provides the protein and coding sequences of human
 CC retinoic acid responsive protein 53.57. The sequences can be used in the
 CC treatment of arrhythmia, bronchial asthma, peptic ulcer, diabetes, cancer
 CC and embryonic development disorders. The present sequence is the protein
 CC of the invention.
 CC
 SQ Sequence 487 AA;
 Query Match 98.8%; Score 1048; DB 23; Length 487;
 Best Local Similarity 99.0%; Pred. No. 2,9e-106;
 Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SQPAGQTSFGATEDYSYGSWYIDEPQGXELQPEGEVPSGHTSTPPGLYHACLASLTL 61
 DB 3 SQPAGQTSFGATEDYSYGSWYIDEPQGXELQPEGEVPSGHTSTPPGLYHACLASLTL 62
 QY 62 VLLLLMLVRRRQLMPDCVGRGRLPSVDFPLAGDRPRAVPAVFWVLLSLCLLLPDED 121
 DB 63 VLLLLMLVRRRQLMPDCVGRGRLPSVDFPLAGDRPRAVPAVFWVLLSLCLLLPDED 122
 QY 122 ALPFLTLASAPSDGKTEAPRGAMKILGLFHYAALYYPLAACATAGHTAAHLIGSTLSWA 181
 DB 123 ALPFLTLASAPSDGKTEAPRGAMKILGLFYAALYYPLAACATAGHTAAHLIGSTLSWA 182
 QY 182 HLGQVQWQRAECQVPKI 199
 DB 183 HLGQVQWQRAECQVPKI 200

RESULT 2
 AAU04557
 ID AAU04557 standard; Protein; 667 AA.
 XX AAU04557;
 AC
 DT 26-SEP-2001 (first entry)
 XX
 XX Human Str26 homologue, PRO10282.
 DE
 XX Human; Str26; PRO10282; 15q23; tumour; breast cancer;
 KW head tumour; leukaemia; lymphoid malignancy; inflammatory disorder;
 KW immunogenic disorder; antigen; antibody.
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT Modified-site 8..12
 FT /note= "Asn is N-glycosylated"
 FT Region 50..56
 FT /label= N_myristoylation_site
 FT Domain 54..69
 FT /label= Transmembrane_domain
 FT Region 89..97
 FT /note= "Region deleted in Str26 variant PRO19578"
 FT Domain 102..119
 FT /label= Transmembrane_domain
 FT Region 132..140
 FT /note= "ATP/GTP-binding site motif A (P-loop)"
 FT Domain 148..166
 FT /label= Transmembrane_domain
 FT Region 176..182
 FT /label= N_myristoylation_site
 FT Domain 207..222
 FT /label= Transmembrane_domain
 FT Region 241..247
 FT /label= N_myristoylation_site
 FT Domain 301..320
 FT /label= Transmembrane_domain

FT Region 317..323
 FT /label= N_myristoylation_site
 FT Region 341..347
 FT /label= N_myristoylation_site
 FT Region 364..375
 FT /note= "Prokaryotic membrane lipoprotein attachment
 FT site"
 FT Domain 364..380
 FT /label= Transmembrane_domain
 FT Domain 431..451
 FT /label= Transmembrane_domain
 FT Domain 474..489
 FT /label= Transmembrane_domain
 FT Region 525..531
 FT /label= N_myristoylation_site
 FT Domain 560..535
 FT /label= Transmembrane_domain
 FT Region 627..633
 FT /label= N_myristoylation_site
 FT Region 631..637
 FT /label= N_myristoylation_site
 FT Region 640..646
 FT /label= N_myristoylation_site
 FT Region 661..667
 FT /label= N_myristoylation_site

MO200151635-A2.
 PD 19-JUL-2001.
 XX
 XX 11-JAN-2001; 2001MO-US00847.
 PF
 XX 13-JAN-2000; 2000US-0175849.
 PR 14-APR-2000; 2000US-0197089.
 PR 29-AUG-2000; 2000US-0228914.
 XX
 XX (GETH) GENENTECH INC.
 PA
 XX Pennica D, Smith V, Wood WJ;
 PI WPI, 2001-442146/47.
 DR N-PSDB; AAS08630.
 DR
 XX Human nucleic acid encoding a PRO10282 or PRO19578 polypeptide (the PRO
 PT polypeptides are referred to as Str26 polypeptides), useful in
 PT molecular biology, including use as hybridisation probes, and in
 PT chromosome and gene mapping -
 XX
 PS Claim 24; Fig 2; 159pp; English.

XX The sequence is human PRO10282, a homologue of murine Str26, a
 CC retinoic acid responsive protein thought to play an important role in
 CC early dorsoventral limb patterning during development and later in the
 CC control of endochondral ossification. The gene for the Str26 homologue
 CC is located on chromosome 15q23. The Str26 polypeptides, agonists,
 CC antagonists or anti-Str26 antibodies are useful for preparation of a
 CC medicament useful in the treatment of a condition which is responsive to
 CC the Str26 polypeptides, agonists, antagonists or anti-Str26 antibodies.
 CC The Str26 polypeptides may also be employed as molecular weight markers
 CC for protein electrophoresis. The Str26 nucleic acids have applications in
 CC molecular biology, including use as hybridisation probes, and in
 CC chromosome and gene mapping. The antibodies and other anti-tumour
 CC compounds may be used to treat various conditions, including those
 CC characterised by overexpression and/or activation of amplified genes.
 CC Exemplary conditions or disorders include benign or malignant tumours
 CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
 CC colorectal, prostate, pancreatic, lung, vulval, thyroid, hepatic
 CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours,
 CC leukemias and lymphoid malignancies), neuronal, glial, astrocytal,
 CC hypochalamic and other glandular, macrophagal, epithelial, stromal and
 CC blastocoele disorders, inflammatory, angiogenic and immunogenic
 CC disorders.
 XX

Sequence 667 AA:

Query Match 98.8%; Score 1048; DB 22; Length 667;
 Best Local Similarity 99.0%; Pred. No. 4.3e-106;
 Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SQPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSPGLYHACLASISIL 61
 DB 3 SQPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSPGLYHACLASISIL 62
 QY 62 VLLLLAMLVRRRLMPDVCVRGPGLPSPVDPLAGDRPRAVPAVFMVLLSSLCULLPED 121
 DB 63 VLLLLAMLVRRRLMPDVCVRGPGLPSPVDPLAGDRPRAVPAVFMVLLSSLCULLPED 122
 QY 122 ALPFLTLASAPSDQKTEAPRGAMKILGLFHYAALYYPLAACATAGHTAAHLGSTLSMA 181
 DB 123 ALPFLTLASAPSDQKTEAPRGAMKILGLFHYAALYYPLAACATAGHTAAHLGSTLSMA 182
 QY 182 HLGQVWQORACCPQVPKI 199
 DB 183 HLGQVWQORACCPQVPKI 200

RESULT 3
 AAB88572
 ID AAB88572 standard; Protein; 667 AA.
 AC AAB88572;
 XX 04-JUN-2001 (first entry)
 DE Human hydrophobic domain containing protein clone HP10713 #36.
 XX Human; hydrophobic domain; immunosuppressant; anti-HIV; neuroprotective;
 KM antianemic; gene therapy; autolimmune disorder; multiple sclerosis;
 KM cytoarrestic; gene therapy; autolimmune disorder; multiple sclerosis;
 KM HIV infection; anaemia; burn; ulcer; osteoporosis; tumour; wound healing;
 KM inflammatory bowel disease; nutritional supplement; appetite; vaccine;
 KM behavioural characteristic; immune response.
 XX Homo sapiens.
 XX MO200112660-A2.
 XX 22-FEB-2001.
 PD 10-AUG-2000; 2000WC-JP05356.
 PF 17-AUG-1999; 99JP-0230344.
 PR 07-SEP-1999; 99JP-0252551.
 PR 01-OCT-1999; 99JP-0281132.
 PR 22-OCT-1999; 99JP-0301624.
 PR 04-NOV-1999; 99JP-0313877.
 PA (SAGA) SAGAMI CHEM RES CENT.
 PA (PROT-) PROTEGENE INC.
 XX Kato S, Kimura T;
 PI WPI: 2001-160059/16.
 DR N-PSDB; AAF94442.
 XX Human proteins with hydrophobic domains and the DNAs which encode them
 PT are useful for treating autoimmune disorders, burns and tumors and for
 PT screening novel pharmaceuticals -
 XX Claim 1; Page 288-291; Slpp; English.
 PS AAF94417 to AAF94516 encode the human proteins given in AAB88572 to
 CC AAB88606 (I) which have a hydrophobic domain. (I) have immunosuppressant,
 CC anti-HIV, neuroprotective, antianemic, vulnary, antilicer,
 CC osteopathic, anti-inflammatory and cyostatic activites, and can be
 CC used in gene therapy. (I) can be used as pharmaceuticals and as antigens

to prepare antibodies. DNA and cDNA (II) encoding (I) can be used as
 CC probes for genetic diagnosis and gene sources for gene therapy or for
 CC producing (I) in large quantities. Cells containing (II) are used for
 CC the detection of ligands or receptors corresponding to membrane or
 CC secretory proteins and to screen small molecule novel pharmaceuticals.
 CC Antibodies directed to (I) can be used for the detection, quantification
 CC and purification of (I). Activities of (I) may include cytokine and cell
 CC proliferation/differentiation function, immune stimulating or suppressing
 CC activity, haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity and anti-inflammatory
 CC activity. (I) and (II) can be used to treat autoimmune disorders e.g.
 CC multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,
 CC inflammatory bowel disease and tumours. (I) and (II) can also be used for
 CC wound healing, as nutritional sources or supplements e.g. as amino acid,
 CC carbon or nitrogen source, to effect metabolism, catabolism, anabolism,
 CC processing and utilisation of dietary fat, protein, carbohydrate, to affect
 CC vitamins and minerals, to effect behavioural characteristics, to affect
 CC appetite, and can act as antigens in vaccines to raise an immune response
 CC to the protein or another material cross-reactive with the protein.
 XX

Sequence 667 AA:

Query Match 98.8%; Score 1048; DB 22; Length 667;
 Best Local Similarity 99.0%; Pred. No. 4.3e-106;
 Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SQPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSPGLYHACLASISIL 61
 DB 3 SQPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSPGLYHACLASISIL 62
 QY 62 VLLLLAMLVRRRLMPDVCVRGPGLPSPVDPLAGDRPRAVPAVFMVLLSSLCULLPED 121
 DB 63 VLLLLAMLVRRRLMPDVCVRGPGLPSPVDPLAGDRPRAVPAVFMVLLSSLCULLPED 122
 QY 122 ALPFLTLASAPSDQKTEAPRGAMKILGLFHYAALYYPLAACATAGHTAAHLGSTLSMA 181
 DB 123 ALPFLTLASAPSDQKTEAPRGAMKILGLFHYAALYYPLAACATAGHTAAHLGSTLSMA 182
 QY 182 HLGQVWQORACCPQVPKI 199
 DB 183 HLGQVWQORACCPQVPKI 200

RESULT 4
 AAU78575
 ID AAU78575 standard; Protein; 667 AA.
 AC AAU78575;
 XX 18-JUN-2002 (first entry)
 DE Human StrA6 (PRO10282) protein.
 XX Human; cancer; StrA6; PRO10282; cyostatic; stromal disorder;
 KM tumour; retinoid; Wnt; tumour antigen; leukaemia; lymphoid malignancy;
 KM neuronal disorder; gliar disorder; astrocytal disorder;
 KM hypochalamic disorder; glandular disorder; macrophagal disorder;
 KM epithelial disorder; blastocoeleic disorder; chromosome 15q23.
 XX Homo sapiens.
 OS
 FH Key
 FT Region
 FT Region
 FT Region
 FT Domain
 FT Domain
 FT Domain
 FT Region

Location/Qualifiers
 8..12
 /note= "N Glycosylation site"
 50..56
 /note= "N-myristoylation site"
 54..69
 /note= "Transmembrane domain"
 102..119
 /note= "Transmembrane domain"
 132..140
 /note= "ATP/GTP binding site motif A"

FT	Domain	148..166
FT		/note="Transmembrane domain"
FT	Region	176..182
FT		/note="N-myristoylation site"
FT	Domain	207..222
FT		/note="Transmembrane domain"
FT	Region	241..247
FT		/note="N-myristoylation site"
FT	Domain	301..320
FT		/note="Transmembrane domain"
FT	Region	317..323
FT		/note="N-myristoylation site"
FT	Region	341..347
FT		/note="N-myristoylation site"
FT	Domain	364..380
FT		/note="Transmembrane domain"
FT	Region	364..375
FT		/note="11poprotein lipid attachment site"
FT	Domain	431..451
FT		/note="Transmembrane domain"
FT	Domain	474..489
FT		/note="Transmembrane domain"
FT	Region	525..531
FT		/note="N-myristoylation site"
FT	Domain	560..535
FT		/note="Transmembrane domain"
FT	Region	627..633
FT		/note="N-myristoylation site"
FT	Region	631..637
FT		/note="N-myristoylation site"
FT	Region	640..646
FT		/note="N-myristoylation site"
FT	Region	661..667
FT		/note="N-myristoylation site"

EN	WO2002186608-A2.
XX	
PD	07-MAR-2002.
XX	
PE	10-JUL-2001; 2001WO-US21635.
XX	
PR	29-AUG-2000; 2000US-228914P.
PR	11-JAN-2001; 2001US-0759056.
XX	
XX	10-JUL-2001; 2001US-0901812.
PA	(GENTH) GENENTECH INC.
PI	
PI	Tice D, Pennica D, Polakis P, Szeto W;
XX	
DR	WPI; 2002-292202/33.
DR	N-P9DB; ABK47278.
XX	
PT	Selectively enhancing the expression of a protein in a tumour cell
PT	characterised by aberrant Wnt signalling, useful for enhancing the
PT	efficacy of treatment of cancers, comprises treating the tumour cell
PT	with a retinoid -
XX	
XX	
PS	Example 1; Fig 2; 141p; English.
XX	
CC	This invention relates to a novel method for selective enhancement of
CC	the expression of a protein in a tumour cell characterised by aberrant
CC	Wnt signalling by treating the tumour cell with a retinoid. The method
CC	is useful for enhancing the efficacy of treatment of cancers
CC	characterised by aberrant Wnt signalling, and for identifying tumour
CC	antigens the expression of which is selectively enhanced by retinoid
CC	treatment. The method is also useful in the early detection of
CC	tumours for early intervention. The tumour antigens are useful as
CC	targets for cancer therapy, since selective enhancement of their
CC	expression by retinoid treatment relative to normal cells improves the
CC	efficacy and therapeutic index of cancer therapeutics directed against
CC	cancer antigens. Antibodies binding to the tumour antigens may be
CC	administered for the treatment of various disorders or conditions,
CC	including those characterised by overexpression and/or activation, of the

CC	tumour antigens, where such conditions or disorders include benign or malignant tumours; leukaemias and lymphoid malignancies; neuronal glial, astrocytal, hypothalamic and other glandular, macrophagal, epithelial, stromal and blastocoealic disorders. The present sequence represents the human Strag protein (PRO10282) which is selectively upregulated by CC reinold using the method of the invention. The gene encoding the human Strag6 protein is located on human chromosome 15q23.
CC	
CC	
CC	
XX	Sequence 667 AA;
SQ	
Query Match	98.8%; Score 1048; DB 23; Length 667;
Best Local Similarity	99.0%; Pred. No. 4.3e-106;
Matches 196; Conservative	1; Mismatches 1; Indels 0; Gaps 0;
OY	2 SOPAGNOTSPGATEDSYSGSWITDEPOCGAKLOPGEVPSCHTSTIPCLYACLASLIL 61
DB	3 SQPANOTSPGATEDSYSGSWITDEPOCGEELOPGEVPSCHTSTIPCLYACLASLIL 62
OY	62 VLLIAMIVRRROLMPDCVRPGLPSPVDLAGRPAVAAPFMYLLSSLCILLPPED 121
DB	63 VLLIAMIVRRROLMPDCVRGRPGLPSPVDLAGRPAVAAPFMYLLSSLCILLPPED 122
OY	122 ALPELTTLASAPSODGKTEAPRGAMKIIGLFHYAALYYPLAACATAGHTAAHLGSTLSMA 181
DB	123 ALPELTTLASAPSODGKTEAPRGAMKIIGLFHYAALYYPLAACATAGHTAAHLGSTLSMA 182
OY	182 HLGQVMQRAECPPVKI 199
DB	183 HLGQVMQRAECPPVKI 200
RESULT 5	
AU083631	
ID	AU083631 standard; Protein; 667 AA.
AC	AU083631;
XX	
DT	08-MAY-2002 (first entry)
XX	
DE	Human PRO protein, Seq ID No 80.
KW	Human; secreted protein; PRO; tumour; lung cancer; colon cancer; breast cancer; prostate tumour; rectal tumour; liver tumour; pericyte cell proliferation; chondrocyte cell proliferation; tumour necrosis factor-alpha.
KM	
KX	
OS	Homo sapiens.
XX	
PN	WO200208288-A2.
PD	
XX	
PF	29-JUN-2001; 2001WO-US21066.
XX	
PR	20-JUL-2000; 2000US-219556P.
PR	25-JUL-2000; 2000US-220585P.
PR	25-JUL-2000; 2000US-220605P.
PR	25-JUL-2000; 2000US-220607P.
PR	25-JUL-2000; 2000US-220624P.
PR	25-JUL-2000; 2000US-220634P.
PR	25-JUL-2000; 2000US-220664P.
PR	25-JUL-2000; 2000US-220666P.
PR	26-JUL-2000; 2000US-220666P.
PR	26-JUL-2000; 2000US-220893B.
PR	28-JUL-2000; 2000WO-US20710.
PR	23-AUG-2000; 2000WO-US2352Z.
PR	24-AUG-2000; 2000WO-US2332B.
PR	15-SEP-2000; 2000US-000000P.
PR	10-NOV-2000; 2000WO-US3087J.
PR	28-NOV-2000; 2000US-253646P.
PR	01-DEC-2000; 2000WO-US3267H.
PR	20-DEC-2000; 2000US-074725S.
PR	20-DEC-2000; 2000WO-US3495C.
PR	28-FEB-2001; 2001WO-US06520.

PR 10-MAY-2001; 2001US-0854280.
 PR 25-MAY-2001; 2001WO-US17092.
 XX
 PA (GENTH) GENENTECH INC.
 XX Baker KP, Desnoyers L, Gerritsen ME, Goddard A, Godowski PJ;
 PI Grimaldi JC, Gurney AL, Smith V, Stephan JF, Watanabe CK, Wood WI;
 XX WPI; 2002-172001/22.
 DR N-PSDB; ABK3575.
 XX
 PT One hundred and twenty two nucleic acids encoding PRO polypeptides,
 PT useful for treating a PRO related disorder and for diagnosing tumours
 PT such as lung cancer, colon cancer, breast tumour, prostate tumour, rectal
 PT tumour or liver tumour -
 PS
 PS Claim 11; Figure 80; 359pp; English.
 XX
 CC The invention relates to one hundred and twenty two nucleic acids
 CC encoding PRO polypeptides. The sequences of the 122 PRO polynucleotides
 CC encode human secreted proteins. The PRO nucleic acids, polypeptides,
 CC agonists and antagonists are useful for treating a PRO related disorder.
 CC The PRO polypeptides are useful for diagnosing tumours, especially lung
 CC cancer, colon cancer, breast tumour, prostate tumour, rectal tumour or
 CC liver tumour. The PRO polypeptides are useful for stimulating the
 CC proliferation of, or gene expression, in pericyte cells, for stimulating
 CC the proliferation or differentiation of chondrocyte cells, for
 CC stimulating the release of tumour necrosis factor-alpha from human blood,
 CC for stimulating or inhibiting the proliferation of normal human dermal
 CC fibroblast cells. The PRO polypeptide may also be used as molecular
 CC weight markers and for tissue typing. The PRO nucleic acids have
 CC applications in molecular biology, including use as hybridisation probes,
 CC and in chromosome and gene mapping. AAU83592-AAU83713 represent human PRO
 CC protein sequences of the invention.
 CC
 XX
 SQ Sequence 667 AA;
 Query Match 98.8%; Score 1048; DB 23; Length 667;
 Best Local Similarity 99.0%; Pred. No. 4.3e-106;
 Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 SQPAGNQTSPGATEDYSGSWYIDEPQGXELQPGCEVPSCHTSIPPGIYHACLASLTL 61
 DB 3 SQPAGNQTSPGATEDYSGSWYIDEPQGXELQPGCEVPSCHTSIPPGIYHACLASLTL 62
 QY 62 VILLIAMLVRRQQLWPCDCYRGRPGLPSPVDFLAGDRPRAVPAVMVLLSSCLLLPDED 121
 DB 63 VILLIAMLVRRQQLWPCDCYRGRPGLPSPVDFLAGDRPRAVPAVMVLLSSCLLLPDED 122
 QY 122 ALPFLTLASAPSDGKTEAPRGAMKILGLFHYAALYYPLAACATAGHTAAHLIGSTLSWA 181
 DB 123 ALPFLTLASAPSDGKTEAPRGAMKILGLFHYAALYYPLAACATAGHTAAHLIGSTLSWA 182
 QY 182 HLGQVQWQRAECPPQVPKI 199
 DB 183 HLGQVQWQRAECPPQVPKI 200
 RESULT 6
 ABB05048
 ID ABB05048 standard; Protein; 667 AA.
 AC ABB05048;
 DT 25-MAR-2002 (first entry)
 XX
 DE Human NOV8c protein SEQ ID NO:32.
 XX
 KM Human; NOVX; cytostatic; antidiabetic; anorectic; antibacterial;
 KM fungicide; virucide; protozoacide; analgesic; antiparkinsonian;
 KM antiaesthetic; hypotensive; osteopathic; antiinflammatory; antitumor;
 KM antiaesthetic; hypotensive; osteopathic; antiinflammatory; antitumor;
 KM neuroprotective; cardiac; antiallergic; antidepressant; nootropic;
 KM anticonvulsant; neuroleptic; gene therapy; vaccine; immune disease;

KM developmental disease; taste and scent detectability disorder; infection;
 KM Burkitt's lymphoma; signal transduction pathway disorder; pain; cancer;
 KM retinal disease; feeding disorder; asthma; Parkinson's disease; ulcer;
 KM noninsulin-dependent diabetes mellitus; acute heart failure; hypotension;
 KM hypertension; urinary retention; osteoporosis; Crohn's disease; allergy;
 KM multiple sclerosis; angina pectoris; myocardial infarction; delirium;
 KM benign prostatic hypertrophy; manic depression; dementia; dyskinesia;
 KM severe mental retardation; Huntington's disease; chromosome 15;
 KM Gilles de la Tourette syndrome.
 OS Homo sapiens.
 XX
 XX W0200190155-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 24-MAY-2001; 2001WO-US17073.
 XX
 PR 24-MAY-2000; 2000US-206679P.
 PR 24-MAY-2000; 2000US-206688P.
 PR 24-MAY-2000; 2000US-206829P.
 PR 30-MAY-2000; 2000US-207748P.
 PR 30-MAY-2000; 2000US-207798P.
 PR 31-MAY-2000; 2000US-208263P.
 PR 02-JUN-2000; 2000US-208831P.
 PR 05-JUN-2000; 2000US-209451P.
 PR 07-JUN-2000; 2000US-210060P.
 PR 20-JUL-2000; 2000US-219507P.
 PR 26-JUL-2000; 2000US-221337P.
 PR 31-JUL-2000; 2000US-221927P.
 PR 19-JAN-2001; 2001US-263135P.
 PR 24-JAN-2001; 2001US-263688P.
 PR 24-JAN-2001; 2001US-263694P.
 PR 23-MAY-2001; 2001US-0863776.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 XX Spytek KA, Majumder K, Tchernev VT, Mishra V, Padigaru M;
 PI Spaderna SK, Shenoy S, Rastelli L, Li L, Taupier RJ, Gangoli E;
 XX WPI; 2002-106174/14.
 DR N-PSDB; ABA92669.
 XX
 PT Novel polypeptide, useful for treating pain, cancer, urinary retention,
 PT osteoporosis, Crohn's disease, diabetes, acute heart failure, dementia,
 PT asthma, ulcer, allergy and Huntington's disease, comprises isolated
 PT polypeptide NOVX -
 XX
 PS Claim 1; Page 106; 266pp; English.
 XX
 CC The present invention describes human NOVX proteins (NOV1 to NOV9). NOVX
 CC proteins (I) have cytostatic, antidiabetic, anorectic, antibacterial,
 CC fungicide, virucide, protozoacide, analgesic, antiparkinsonian,
 CC antiaesthetic, hypotensive, osteopathic, antiinflammatory, antitumor,
 CC neuroprotective, cardiac, antiallergic, antidepressant, nootropic, (II)
 CC anticonvulsant and neuroleptic activities. (I) and polynucleotides (II)
 CC can be used in gene therapy and vaccine production. (I) and (II) can be
 CC used for treating or preventing a NOVX-associated disorder such as
 CC cardiomyopathy, atherosclerosis and diabetes in a human, where the
 CC disorder is related to cell signal processing and metabolic path way
 CC modulation, in a subject, preferably human. (I) and (II) can be used for
 CC diagnosing, preventing or treating developmental diseases, immune
 CC diseases, taste and scent detectability disorder, Burkitt's lymphoma,
 CC signal transduction pathway disorders, retinal diseases including those
 CC involving photoreception, cell growth rate disorders, feeding disorders,
 CC noninsulin-dependent diabetes mellitus, infections, pain, cancer, asthma,
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,
 CC urinary retention, osteoporosis, Crohn's disease, multiple sclerosis,
 CC Albrigth Hereditary Osteodystrophy, angina pectoris, ulcers, myocardial
 CC infarction, allergies, benign prostatic hypertrophy, manic depression,
 CC delirium, dementia, severe mental retardation and dyskinesias, such as
 CC Huntington's disease or Gilles de la Tourette syndrome. The present
 CC sequence represents the human NOV8c protein from the present invention.

CC Human NOV2a is isolated to chromosome 15.
 XX
 XX Sequence 667 AA:
 Query Match 98.7%; Score 1047; DB 23; Length 667;
 Best Local Similarity 98.5%; Pred. No. 5.6e-106;
 Matches 195; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 SOPAGNOTSPGATEDSYGSMYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLITL 61
 DB 3 SQPAGNOTSPGATEDSYGSMYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLITL 62
 QY 62 VLLLLMLVRRROLMPDCVGRPGPLSPVDFLAGDRPRAVPAVFMVLLSSLCLLPDED 121
 DB 63 VLLLLMLVRRROLMPDCVGRPGPLSPVDFLAGDRPRAVPAVFMVLLSSLCLLPDED 122
 QY 122 ALPELTLASAPSDGKTEAPRGAMKILGLFYAALYPLAACATAGHTAAHLGSLTSLMA 181
 DB 123 ALPELTLASAPSDGKTEAPRGAMKILGLFYAALYPLAACATAGHTAAHLGSLTSLMA 182
 QY 182 HLGVOVMQRAECPQVPKI 199
 DB 183 HLGVOVMQRAECPQVPKI 200

RESULT 7
 ABB05047
 ID ABB05047 standard; Protein; 662 AA.
 XX ABB05047;
 XX
 DT 25-MAR-2002 (first entry)
 XX
 DE Human NOV8b protein SEQ ID NO:30.
 XX
 XX Human; NOXV; cytostatic; antidiabetic; anorectic; antibacterial;
 KM fungicide; virucide; protozoacide; analgesic; antiparkinsonian;
 KM antiaesthetic; hypotensive; osteopathic; antiinflammatory; antitumor;
 KM neuroprotective; gardiant; antiallergic; antidepressant; nootropic;
 KM anticonvulsant; neuroleptic; gene therapy; vaccine; immune disease;
 KM developmental disease; taste and scent detectability disorder; infection;
 KM Burkitt's lymphoma; signal transduction pathway disorder; pain; cancer;
 KM retinal disease; feeding disorder; asthma; Parkinson's disease; ulcer;
 KM noninsulin-dependent diabetes mellitus; acute heart failure; hypotension;
 KM hypertension; urinary retention; osteoporosis; Crohn's disease; allergy;
 KM multiple sclerosis; angina pectoris; myocardial infarction; delirium;
 KM benign prostatic hypertrophy; manic depression; dementia; dyskinesia;
 KM severe mental retardation; Huntington's disease;
 KM Gilles de la Tourette syndrome.
 KM
 OS Homo sapiens.
 XX
 XX MO200190155-A2.
 XX
 PD 29-NOV-2001.
 XX
 XX 24-MAY-2001; 2001WO-US17073.
 XX
 XX 24-MAY-2000; 2000US-206679P.
 PR 24-MAY-2000; 2000US-206688P.
 PR 24-MAY-2000; 2000US-206829P.
 PR 30-MAY-2000; 2000US-207748P.
 PR 30-MAY-2000; 2000US-207798P.
 PR 31-MAY-2000; 2000US-208263P.
 PR 02-JUN-2000; 2000US-208831P.
 PR 05-JUN-2000; 2000US-209451P.
 PR 07-JUN-2000; 2000US-210060P.
 PR 20-JUL-2000; 2000US-219507P.
 PR 26-JUL-2000; 2000US-221337P.
 PR 31-JUL-2000; 2000US-221927P.
 PR 19-JAN-2001; 2001US-263135P.
 PR 24-JAN-2001; 2001US-263688P.
 PR 24-JAN-2001; 2001US-263694P.

PR 23-MAY-2001; 2001US-0863776.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 XX Spytek KA, Majumder K, Tchernev VT, Mishra V, Padigaru M,
 PI Spaderna SK, Shenoy S, Rastelli L, Li L, Taupier RJ, Gangoli E,
 XX
 DR WPI; 2002-106174/14.
 DR N-PsDB; ABA92666.
 XX
 PT Novel polypeptide, useful for treating pain, cancer, urinary retention,
 PT osteoporosis, Crohn's disease, diabetes, acute heart failure, dementia,
 PT asthma, ulcer, allergy and Huntington's disease, comprises isolated
 PT polypeptide NOVX -
 XX
 PS Claim 1; Page 103; 26pp; English.
 XX
 CC The present invention describes human NOVX proteins (NOV1 to NOV9). NOVX
 CC proteins (I) have cytostatic, antidiabetic, anorectic, antibacterial,
 CC fungicide, virucide, protozoacide, analgesic, antiparkinsonian,
 CC antiaesthetic, hypotensive, osteopathic, antiinflammatory, antitumor,
 CC neuroprotective, gardiant, antiallergic, antidepressant, nootropic,
 CC anticonvulsant and neuroleptic activities. (I) and polynucleotides (II)
 CC can be used in gene therapy and vaccine production. (I) and (II) can be
 CC used for treating or preventing a NOVX-associated disorder such as
 CC cardiomyopathy, atherosclerosis and diabetes in a human, where the
 CC disorder is related to cell signal processing and metabolic path way
 CC modulation, in a subject, preferably human. (I) and (II) can be used for
 CC diagnosing, preventing or treating developmental diseases, immune
 CC diseases, taste and scent detectability disorder, Burkitt's lymphoma,
 CC signal transduction pathway disorders, retinal diseases including those
 CC involving photoreception, cell growth rate disorders, feeding disorders,
 CC noninsulin-dependent diabetes mellitus, infections, pain, cancer, asthma,
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,
 CC urinary retention, osteoporosis, Crohn's disease, multiple sclerosis,
 CC Albrigth Hereditary Osteodystrophy, angina pectoris, ulcers, myocardial
 CC infarction, allergies, benign prostatic hypertrophy, manic depression,
 CC delirium, dementia, severe mental retardation and dyskinesias, such as
 CC Huntington's disease or Gilles de la Tourette syndrome. The present
 CC sequence represents the human NOV8b protein from the present invention.
 XX
 SQ Sequence 662 AA:
 Query Match 98.3%; Score 1043; DB 23; Length 662;
 Best Local Similarity 98.5%; Pred. No. 1.5e-105;
 Matches 195; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 SOPAGNOTSPGATEDSYGSMYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLITL 61
 DB 3 SQPAGNOTSPGATEDSYGSMYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLITL 62
 QY 62 VLLLLMLVRRROLMPDCVGRPGPLSPVDFLAGDRPRAVPAVFMVLLSSLCLLPDED 121
 DB 63 VLLLLMLVRRROLMPDCVGRPGPLSPVDFLAGDRPRAVPAVFMVLLSSLCLLPDED 122
 QY 122 ALPELTLASAPSDGKTEAPRGAMKILGLFYAALYPLAACATAGHTAAHLGSLTSLMA 181
 DB 123 ALPELTLASAPSDGKTEAPRGAMKILGLFYAALYPLAACATAGHTAAHLGSLTSLMA 182
 QY 182 HLGVOVMQRAECPQVPKI 199
 DB 183 HLGVOVMQRAECPQVPKI 200

RESULT 8
 AAU91278
 ID AAU91278 standard; Protein; 689 AA.
 XX AAU91278;
 AC
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Human NOV2b protein.

XX Human; NOVX; gene therapy; cardiomyopathy; atherosclerosis;
 KW diabetes; cell signal processing; metabolic pathway modulation;
 KW inflammation; autoimmune disorder; scleroderma; transplantation;
 KW allergy; systemic lupus erythematosus; haemophilia; Alzheimer's disease;
 KW graft versus host disease; Leech-Nyhan syndrome; periodontitis;
 KW pancreatitis; musculoskeletal disorder; Parkinson's disease;
 KW Huntington's disease; behavioural disorder; pain; obesity; wound healing;
 KW neurodegenerative disorder; neuropsychiatric disorder; hypertension;
 KW growth disorder; reproductive disorder; lung disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200216600-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 27-AUG-2001; 2001MO-US26518.
 XX
 PR 25-AUG-2000; 2000US-227800P.
 PR 25-AUG-2000; 2000US-228205P.
 PR 25-AUG-2000; 2000US-228324P.
 PR 30-AUG-2000; 2000US-228997P.
 PR 30-AUG-2000; 2000US-228185P.
 PR 01-SEP-2000; 2000US-228780P.
 PR 01-SEP-2000; 2000US-229848P.
 PR 01-SEP-2000; 2000US-229850P.
 PR 22-JAN-2001; 2001US-26337P.
 PR 31-JAN-2001; 2001US-265518P.
 PR 15-MAR-2001; 2001US-276451P.
 PR 27-MAR-2001; 2001US-279196P.
 PR 24-AUG-2001; 2001US-0393398.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Gerlach V, MacDougall JR, Smithson G, Stone DJ, Ellerman K;
 PI Spytek KA, Zernusen BD, Rastelli L, Verney CAM, Paturajan M;
 PI Tchiernev VT, Padigaru M, Taupier RJ;
 XX
 DR WPI; 2002-292064/33.
 DR N-PSDB; ABK55562.
 XX
 PT New isolated cytoplasmic, nuclear, membrane bound and secreted
 PT polypeptides, termed NOVX, useful for treating inflammation, autoimmune
 PT disorders, haemophilia, Leech-Nyhan syndrome, pancreatitis,
 PT musculoskeletal disorders -
 PT
 PS Claim 1; Page 20; 245pp; English.
 XX
 CC The invention relates to an isolated cytoplasmic, nuclear, membrane bound
 CC or secreted polypeptide, designated NOVX (actually NOV1, 2a, 2b, 3a, 3b,
 CC 4, 5a, 5b, 5c, 5d, 5e, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOVX, a
 CC mature form, or a variant of the mature form of NOVX. Also included
 CC are a polynucleotide encoding NOVX (or its complement), a vector
 CC comprising the polynucleotide, a cell comprising the vector, an
 CC anti-NOVX antibody, determining the presence of NOVX in a sample
 CC using the antibody, determining the presence of NOVX polynucleotide in
 CC a sample using a probe which binds to NOVX polynucleotide, identifying a
 CC an agent which binds to NOVX (including modulators of NOVX). NOVX, the
 CC polynucleotide and the antibody are useful for diagnosing, treating
 CC or preventing a NOVX-associated disorder selected from cardiomyopathy,
 CC atherosclerosis, diabetes, a disorder related to cell signal processing
 CC and metabolic pathway modulation, inflammation, autoimmune disorders,
 CC scleroderma, transplantation, allergies, systemic lupus erythematosus,
 CC haemophilia, graft versus host disease, Alzheimer's disease, stroke,
 CC Leesch-Nyhan syndrome, periodontitis, pancreatitis, musculoskeletal
 CC disorders, Parkinson's disease, Huntington's disease, behavioural
 CC disorders, pain, neurodegenerative and neuropsychiatric disorders,
 CC hyperension, wound healing, obesity, growth and reproductive
 CC disorders, lung diseases and many other diseases and disorders listed in
 CC the specification. NOVX, the polynucleotide and the antibody are useful
 CC in screening assays, detection assays (e.g., chromosomal mapping, tissue
 CC typing, forensic biology), predictive medicine (e.g., diagnostic assays,

CC prognostic assays, monitoring clinical trials and pharmacogenomic), and
 CC in methods of treatment (e.g., therapeutic and prophylactic). NOVX is
 CC useful as immunogen to produce antibodies immunospecific for NOVX, as
 CC vaccines to screen for potential agonist and antagonist compounds, and
 CC as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide
 CC is useful in gene therapy, to express NOVX, to detect NOVX mRNA
 CC or a genetic lesion in a NOVX gene, and to modulate NOVX activity. The
 CC vector is useful for producing non-human transgenic animals. The antibody
 CC is useful for isolating, and purifying NOVX and to monitor protein levels
 CC in tissue as part of a clinical testing procedure. The present sequence
 CC represents a NOVX protein.
 XX
 SO Sequence 689 AA;
 XX
 Query Match 96.1%; Score 1020; DB 23; Length 689;
 Best Local Similarity 88.6%; Pred. No. 5.3e-103;
 Matches 195; Conservative 1; Mismatches 2; Indels 22; Gaps 1;
 QY 2 SQPAGNQTSPGATEDYSGSWYIDEPQGXLOPBGEPSPCHTSPGLYHACLASISIL 61
 DB 3 SQPAGNQTSPGATEDYSGSWYIDEPQGXLOPBGEPSPCHTSPGLYHACLASISIL 62
 QY 62 VLLLLAMLVRRRLMPCCVRGRPGLP-----SPVDPLAGDRPR 99
 DB 63 VLLLLAMLVRRRLMPCCVRGRPGLPSPSEWNGVQGLSNSLLPCCLCASPVDFLAGDPR 122
 QY 100 AVPAVFWVLSSICLLPDEDALPFLTLASPSQDGKTEAPRGAKWILGLFHYAALYYP 159
 DB 123 AVPAVFWVLSSICLLPDEDALPFLTLASPSQDGKTEAPRGAKWILGLFHYAALYYP 182
 QY 160 LAACATGHTAAHLGLSTLSAAHLGVOYMGRAECPOVPKI 199
 DB 183 LAACATGHTAAHLGLSTLSAAHLGVOYMGRAECPOVPKI 222
 XX
 RESULT 9
 AAU04558
 ID AAU04558 standard; Protein: 658 AA.
 XX
 AC AAU04558;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE Human Str66 homologue, PRO19578.
 XX
 KW Human; Str66; PRO19578; 15q23; tumour; breast cancer;
 KW head tumour; leukaemia; lymphoid malignancy; inflammatory disorder;
 KW immunogenic disorder; antigen; antibody.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 8..12 /note= "Asn is N-glycosylated"
 FT Region 50..56 /label= N_myristoylation_site
 FT Domain 54..71 /label= Transmembrane_domain
 FT Domain 93..111 /label= Transmembrane_domain
 FT Region 123..131 /note= "ATP/GTP-binding site motif A (P-loop)"
 FT Domain 140..157 /label= Transmembrane_domain
 FT Region 167..173 /label= N_myristoylation_site
 FT Domain 197..214 /label= Transmembrane_domain
 FT Region 232..238 /label= N_myristoylation_site
 FT Domain 291..312 /label= Transmembrane_domain
 FT Region 308..314

FT	/label=N_myristoylation_site
FT	332..338
FT	/label=N_myristoylation_site
FT	355..366
FT	/note="Prokaryotic membrane lipoprotein attachment
FT	site"
FT	356..371
FT	/label=Transmembrane_domain
FT	425..444
FT	/label=Transmembrane_domain
FT	464..481
FT	/label=Transmembrane_domain
FT	505..522
FT	/label=Transmembrane_domain
FT	516..522
FT	/label=N_myristoylation_site
FT	618..624
FT	/label=N_myristoylation_site
FT	622..628
FT	/label=N_myristoylation_site
FT	631..637
FT	/label=N_myristoylation_site
FT	652..658
FT	/label=N_myristoylation_site
XX	
PN	WO200151635-A2.
PD	
XX	
PN	19-JUL-2001.
XX	
PN	11-JAN-2001; 2001WO-US00847.
XX	
PR	13-JAN-2000; 2000US-0175849.
PR	14-APR-2000; 2000US-0197089.
PR	29-AUG-2000; 2000US-0228914.
XX	
PA	(GETH) GENENTECH INC.
PI	
PI	Pennica D, Smith V, Wood WI;
XX	
DR	WPI; 2001-442146/47.
DR	N-PEDB; AAS08632.
XX	
PT	Human nucleic acid encoding a PRO10282 or PRO19578 polypeptide (the PRO
PT	polypeptides are referred to as Stra6 polypeptides), useful in
PT	molecular biology, including use as hybridisation probes, and in
PT	chromosome and gene mapping -
XX	
PS	Claim 24; Fig 7; 15pp; English.
XX	
CC	The sequence is human PRO19578 (a variant possibly due to
CC	alternative splicing of the Stra6 mRNA), a homologue of murine Stra6, a
CC	retinoic acid responsive protein thought to play an important role in
CC	early dorsoventral limb patterning during development and later in the
CC	control of endochondral ossification. The gene for the Stra6 homologue
CC	is located on chromosome 15q23. The Stra6 polypeptides, agonists,
CC	antagonists or anti-Stra6 antibodies are useful for preparation of a
CC	medicament useful in the treatment of a condition which is responsive to
CC	the Stra6 polypeptides, agonists, antagonists or anti-Stra6 antibodies.
CC	The Stra6 polypeptides may also be employed as molecular weight markers
CC	for protein electrophoresis. The Stra6 nucleic acids have applications in
CC	molecular biology, including use as hybridisation probes, and in
CC	chromosome and gene mapping. The antibodies and other anti-tumour
CC	compounds may be used to treat various conditions, including those
CC	characterised by overexpression and/or activation of amplified genes.
CC	Exemplary conditions or disorders include benign or malignant tumours
CC	(e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
CC	colorectal, prostate, pancreatic, lung, vulvar, thyroid, hepatic
CC	carcinomas, sarcomas, glioblastomas, and various head and neck tumours,
CC	leukaemias and lymphoid malignancies), neuronal, glial, astrocytal,
CC	hypothalamic and other glandular, macrophagal, epithelial, stromal and
CC	haematocell disorders, inflammatory, angiogenic and immunogenic
CC	disorders.
XX	

Seq	Sequence	658 AA;	93.0%; Score 986.5; DB 22; Length 658; Beat Local Similarity 94.4%; Pred. No. 2.4e-99; Matches 187; Conservative 1; Mismatches 1; Indels 9; Gaps 1;
QY	2	SOPANQNTSPQATEEYISGSGWIDEPQGXELQEPGEIPSCHTSIPPLLYACLASLSIL	61
Db	3	SOPANQNTSPGTEDEYSXGSWIDEPQGEELQEPGEVPSCHTSIPPLLYACLASLSIL	62
QY	62	VLLIAMLYRRRQLMPCDVRGRPGLPSPVDVLADGRPAVAAYFMVLLSLCLLPDED	121
Db	63	VLLIAMLYRRRQLMPCDVRGRPGLP-----RPAAYPAAYFMVLLSLCLLPDED	113
QY	122	ALPFLTLASABSQDCKTEAPRGAMKILGLPHYALYYPLAACATAGHTAAHLGSLTSLWA	181
Db	114	ALPFLTLASAPSDQCKTEAPRGAMKILGLFYALYYPLAACATAGHTAAHLGSLTSLWA	173
QY	182	HLGVQVMQRAECPOVPKI	199
Db	174	HLGVQVMQRAECPOVPKI	191
RESULT 10			
AAU78576	AAU78576 standard; Protein; 658 AA.		
XX	AAU78576;		
XX	18-JUN-2002	(first entry)	
DE	Human Str66 protein variant (PRO19578).		
KW	Human; cancer; Str66; PRO19578; cytostatic; stromal disorder;		
KM	tumour; retinoid; Mrt; tumour antigen; leukaemia; lymphoid malignancy;		
KM	neural disorder; gliol disorder; astrocytal disorder;		
KW	hypothalamic disorder; glandular disorder; macophagal disorder;		
KW	epithelial disorder; blascoceolic disorder; chromosome 15q23.		
OS	Homo sapiens.		
XX			
XX			
PH	Key	Location/Qualifiers	
FT	Region	8..12 /note= "N Glycosylation site"	
FT	Region	50..56 /note= "N-myristoylation site"	
FT	Domain	54..71 /note= "Transmembrane domain"	
FT	Domain	93..111 /note= "Transmembrane domain"	
FT	Region	123..131 /note= "ATP/GTP binding site motif A"	
FT	Domain	140..157 /note= "Transmembrane domain"	
FT	Region	167..173 /note= "N-myristoylation site"	
FT	Domain	197..214 /note= "Transmembrane domain"	
FT	Region	232..238 /note= "N-myristoylation site"	
FT	Domain	291..312 /note= "Transmembrane domain"	
FT	Region	308..314 /note= "N-myristoylation site"	
FT	Region	332..338 /note= "N-myristoylation site"	
FT	Domain	356..371 /note= "Transmembrane domain"	
FT	Region	355..366 /note= "lipoprotein lipid attachment site"	
FT	Domain	425..444 /note= "Transmembrane domain"	
FT	Domain	464..481	

```

FT      Region      /note= "Transmembrane domain"
FT      Region      516..522
FT      Domain      /note= "N-myristoylation site"
FT      Domain      505..522
FT      Region      /note= "Transmembrane domain"
FT      Region      618..628
FT      Region      /note= "N-myristoylation site"
FT      Region      622..628
FT      Region      /note= "N-myristoylation site"
FT      Region      631..637
FT      Region      /note= "N-myristoylation site"
FT      Region      652..658
FT      Region      /note= "N-myristoylation site"
XX      WO200218608-A2.
XX      PD      07-MAR-2002.
XX      PF      10-JUL-2001; 2001WO-US21635.
XX      PR      29-AUG-2000; 2000US-228914P.
XX      PR      11-JAN-2001; 2001US-0759056.
XX      PR      10-JUN-2001; 2001US-0901812.
XX      PA      (GETH ) GENENTECH INC.
XX      PI      Tice D, Pennica D, Polakis P, Szeto W;
XX      DR      WPI; 2002-292202/33.
XX      DR      N-PSDB; ABK47280.
XX      PT      Selectively enhancing the expression of a protein in a tumour cell
XX      PT      characterised by aberrant Wnt signalling, useful for enhancing the
XX      PT      efficacy of treatment of cancers, comprises treating the tumour cell
XX      PT      with a retinoid
XX      Example 1; Fig 7; 141pp, English.
XX      CC      This invention relates to a novel method for selective enhancement of
XX      CC      the expression of a protein in a tumour cell characterised by aberrant
XX      CC      Wnt signalling by treating the tumour cell with a retinoid. The method
XX      CC      is useful for enhancing the efficacy of treatment of cancers
XX      CC      characterised by aberrant Wnt signalling, and for identifying tumour
XX      CC      antigens the expression of which is selectively enhanced by retinoid
XX      CC      treatment. The method is also useful in the early detection of
XX      CC      tumours for early intervention. The tumour antigens are useful as
XX      CC      targets for cancer therapy, since selective enhancement of their
XX      CC      expression by retinoid treatment relative to normal cells improves the
XX      CC      efficacy and therapeutic index of cancer therapeutics directed against
XX      CC      these antigens. Antibodies binding to the tumour antigens may be
XX      CC      administered for the treatment of various disorders or conditions,
XX      CC      including those characterised by overexpression and/or activation of the
XX      CC      tumour antigens, where such conditions or disorders include benign or
XX      CC      malignant tumours; leukaemias and lymphoid malignancies; neuronal, glial,
XX      CC      astrocytal, hypothalamic and other glandular, macrophagal, epithelial,
XX      CC      stromal and blastocoealic disorders. The present sequence represents
XX      CC      the human Stra6 protein variant (PRO19578) which is actively upregulated
XX      CC      by retinoid using the method of the invention. The gene encoding the
XX      CC      human Stra6 protein is located on human chromosome 15q23.
XX      SQ      Sequence 658 AA;
XX      Query Match 93.0%; Score 986.5; DB 23; Length 658;
XX      Best Local Similarity 94.4%; Pred. No. 2.4e-99;
XX      Matches 187; Conservative 1; Mismatches 1; Indels 9; Gaps 1;
XX      QY      2 SQPAGNQTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSPGLYNACLASLSIL 61
XX      DB      3 SQPAGNQTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSPGLYNACLASLSIL 62
XX      QY      62 VLLLLAMLVRRRLQMPDCVRRGRPGILPSPVDFLAGRPPAIVPAVWVLLSSICLLLPED 121
XX      DB      63 VLLLLAMLVRRRLQMPDCVRRGRPGILPSPVDFLAGRPPAIVPAVWVLLSSICLLLPED 113

```

```

QY      122 ALPFLTLASAPSDQKTEAPRGAMKILGLFYFVYALYPLAACATAGHTAAILGSTLSMA 181
DB      114 ALPFLTLASAPSDQKTEAPRGAMKILGLFYFVYALYPLAACATAGHTAAILGSTLSMA 173
QY      182 HLGQVWQRAECPOVPKI 199
DB      174 HLGQVWQRAECPOVPKI 191
RESULT 11
ABB05046
ID      ABB05046 standard; Protein; 653 AA.
XX      AC      ABB05046;
XX      DT      25-MAR-2002 (first entry)
XX      DE      Human NOV8a protein SEQ ID NO:28.
XX      KW      Human; NOX; cytosstatic; antidiabetic; anorectic; antibacterial;
XX      KW      fungicide; virucide; protozoacide; analgesic; antiparkinsonian;
XX      KW      antiaesthetic; hypotensive; osteopathic; antiinflammatory; antiulcer;
XX      KW      neuroprotective; cardiac; antiallergic; antidepressant; nootropic;
XX      KW      anticonvulsant; neuroleptic; gene therapy; vaccine; immune disease;
XX      KW      developmental disease; taste and scent detectability disorder; infection;
XX      KW      Burkitt's lymphoma; signal transduction pathway disorder; pain; cancer;
XX      KW      retinal disease; feeding disorder; asthma; Parkinson's disease; ulcer;
XX      KW      noninsulin-dependent diabetes mellitus; acute heart failure; hypotension;
XX      KW      hypertension; urinary retention; osteoporosis; Crohn's disease; allergy;
XX      KW      multiple sclerosis; angina pectoris; myocardial infarction; delirium;
XX      KW      benign prostatic hypertrophy; manic depression; dementia; dyskinesia;
XX      KW      severe mental retardation; Huntington's disease; chromosome 15;
XX      KW      Gilles de la Tourette syndrome.
XX      OS      Homo sapiens.
XX      PN      WO200190155-A2.
XX      PD      29-NOV-2001.
XX      PF      24-MAY-2001; 2001WO-US17073.
XX      PR      24-MAY-2000; 2000US-206679P.
XX      PR      24-MAY-2000; 2000US-206688P.
XX      PR      24-MAY-2000; 2000US-206829P.
XX      PR      30-MAY-2000; 2000US-207748P.
XX      PR      30-MAY-2000; 2000US-207798P.
XX      PR      31-MAY-2000; 2000US-208263P.
XX      PR      02-JUN-2000; 2000US-208831P.
XX      PR      05-JUN-2000; 2000US-209451P.
XX      PR      07-JUN-2000; 2000US-210060P.
XX      PR      20-JUL-2000; 2000US-219507P.
XX      PR      26-JUL-2000; 2000US-221337P.
XX      PR      31-JUL-2000; 2000US-221927P.
XX      PR      19-JAN-2001; 2001US-263135P.
XX      PR      24-JAN-2001; 2001US-263688P.
XX      PR      24-JAN-2001; 2001US-263694P.
XX      PR      23-MAY-2001; 2001US-0863776.
XX      PA      (CURA-) CURAGEN CORP.
XX      PI      Spytek Ks, Majumder K, Tchertnev VT, Miehra V, Padigaru M;
XX      PI      Spaderna SK, Shenoy S, Rastelli L, Li L, Taupier RJ, Gangoli E;
XX      WPI; 2002-106174/14.
XX      DR      N-PSDB; ABA92663.
XX      PT      Novel polypeptide, useful for treating pain, cancer, urinary retention,
XX      PT      osteoporosis, Crohn's disease, diabetes, acute heart failure, dementia,
XX      PT      asthma, ulcer, allergy and Huntington's disease, comprises isolated
XX      PT      polypeptide NOV8 -

```

PS Claim 1; Page 98; 266pp; English.

XX The present invention describes human NOVX proteins (NOVX to NOV9). NOVX
 CC proteins (I) have cytostatic, antidiabetic, anorectic, antibacterial,
 CC fungicide, virucide, proteosacide, analgesic, antiparkinsonian,
 CC antiaesthetic, hypotensive, osteopathic, antiinflammatory, antitumor,
 CC neuroprotective, cardiant, antiallergic, antidepressant, nocotropic, (II)
 CC anticonvulsant and neuroleptic activities. (I) and polynucleotides (II)
 CC can be used in gene therapy and vaccine production. (I) and (II) can be
 CC used for treating or preventing a NOVX-associated disorder such as
 CC cardiomyopathy, atherosclerosis and diabetes in a human, where the
 CC disorder is related to cell signal processing and metabolic path way
 CC modulation, in a subject, preferably human. (I) and (II) can be used for
 CC diagnosing, preventing or treating developmental diseases, immune
 CC diseases, taste and scent detectability disorder, Burkitt's lymphoma,
 CC involving photoreception, cell growth rate disorders, feeding disorders,
 CC noninsulin-dependent diabetes mellitus, infections, pain, cancer, asthma,
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,
 CC urinary retention, osteoporosis, Crohn's disease, multiple sclerosis,
 CC Albright Hereditary Osteodystrophy, angina pectoris, ulcers, myocardial
 CC infarction, allergies, benign prostatic hypertrophy, manic depression,
 CC delirium, dementia, severe mental retardation and dyskinesias, such as
 CC Huntington's disease or Gilles de la Tourette syndrome. The present
 CC sequence represents the human NOV8a protein from the present invention.
 CC Human NOV8a is isolated to chromosome 15.

XX Sequence 653 AA:

Query Match 92.9%; Score 985.5; DB 23; Length 653;
 Best Local Similarity 94.4%; Pred. No. 3,1e-99;
 Matches 187; Conservative 1; Mismatches 1; Indels 9; Gaps 1;

QY 2 SQPAAQNSPGATERYSGSWYIDPQGXELQPEGEVSCSTSPGLYHACLASLTL 61
 DB 3 SQPAAQNSPGATERYSGSWYIDPQGXELQPEGEVSCSTSPGLYHACLASLTL 62
 QY 62 VLLIAMIVRRQQLWPDCCVGRGRLPSPPDFLAGDRPRAVPAVFMVLLSCLLPDED 121
 DB 63 VLLIAMIVRRQQLWPDCCVGRGRLPSPPDFLAGDRPRAVPAVFMVLLSCLLPDED 122
 QY 122 ALPFTLLASAPSDQKTEAPRGAMKILGFHYAALYPLAACATAGHTAAHLGSLTSLWA 181
 DB 123 ALPFTLLASAPSDQKTEAPRGAMKILGFHYAALYPLAACATAGHTAAHLGSLTSLWA 173
 QY 182 HLGQVQWQRAECQVPKI 199
 DB 174 HLGQVQWQRAECQVPKI 191

RESULT 12

AAU91277 standard; Protein; 660 AA.

AAU91277;

18-JUN-2002 (first entry)

Human NOV8a protein.

Human; NOVX; gene therapy; cardiomyopathy; atherosclerosis;
 diabetes; cell signal processing; metabolic pathway modulation;
 inflammation; autoimmune disorder; scleroderma; transplantation;
 allergy; systemic lupus erythematosus; haemophilia; Alzheimer's disease;
 graft versus host disease; Leisch-Nyhan syndrome; periodontitis;
 pancreaticitis; musculoskeletal disorder; Parkinson's disease;
 Huntington's disease; behavioural disorder; pain; obesity; wound healing;
 neurodegenerative disorder; neuropsychiatric disorder; hypertension;
 growth disorder; reproductive disorder; lung disease.

Homo sapiens.

WO200216600-A2.

XX 28-FEB-2002.

XX 27-AUG-2001; 2001WO-US66518.

XX 25-AUG-2000; 2000US-227800P.
 XX 25-AUG-2000; 2000US-228205P.
 XX 25-AUG-2000; 2000US-228324P.
 XX 30-AUG-2000; 2000US-228997P.
 XX 30-AUG-2000; 2000US-229185P.
 XX 01-SEP-2000; 2000US-229780P.
 XX 01-SEP-2000; 2000US-229846P.
 XX 01-SEP-2000; 2000US-229850P.
 XX 22-JAN-2001; 2001US-263337P.
 XX 31-JAN-2001; 2001US-265518P.
 XX 15-MAR-2001; 2001US-276451P.
 XX 27-MAR-2001; 2001US-279196P.
 XX 24-AUG-2001; 2001US-0393398.

(CURA-) CURAGEN CORP.

XX Gerlach V, MacDougall JR, Smithson G, Stone DJ, Ellerman K;
 PI Spryck KA, Zernusen BD, Rastelli L, Verney CM, Paturnajan M;
 XX Tchernev VT, Padigaru M, Taupier RJ;

DR WPI; 2002-292064/33.
 DR N-PSDB; ABK55561.

XX New isolated cytoplasmic, nuclear, membrane bound and secreted
 PT polypeptides, termed NOVX, useful for treating inflammation, autoimmune
 PT disorders, haemophilia, Leisch-Nyhan syndrome, pancreaticitis,
 PT musculoskeletal disorders -
 PS Claim 1; Page 18; 245pp; English.

XX The invention relates to an isolated cytoplasmic, nuclear, membrane bound
 CC or secreted polypeptide, designated NOVX (actually NOV1, 2a, 2b, 3a, 3b,
 CC 4, 5a, 5b, 5c, 5d, 5e, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOVX, a
 CC mature form, or a variant of the mature form of NOVX. Also included
 CC are a polynucleotide encoding NOVX (or its complement), a vector
 CC comprising the polynucleotide, a cell comprising the vector, an
 CC anti-NOVX antibody, determining the presence of NOVX in a sample
 CC using the antibody, determining the presence of NOVX polynucleotide in
 CC a sample using a probe which binds to NOVX polynucleotide, identifying a
 CC an agent which binds to NOVX (including modulators of NOVX). NOVX, the
 CC polynucleotide and the antibody are useful for diagnosing, treating
 CC or preventing a NOVX-associated disorder selected from cardiomyopathy,
 CC atherosclerosis, diabetes, a disorder related to cell signal processing,
 CC and metabolic pathway modulation, inflammation, autoimmune disorders,
 CC scleroderma, transplantation, allergies, systemic lupus erythematosus,
 CC haemophilia, graft versus host disease, Alzheimer's disease, stroke,
 CC Leisch-Nyhan syndrome, periodontitis, pancreaticitis, musculoskeletal
 CC disorders, Parkinson's disease, Huntington's disease, behavioural
 CC disorders, pain, neurodegenerative and neuropsychiatric disorders,
 CC hypertension, wound healing, obesity, growth and reproductive
 CC disorders, lung diseases and many other diseases and disorders listed in
 CC the specification. NOVX, the polynucleotide and the antibody are useful
 CC in screening assays, detection assays (e.g., chromosomal mapping, tissue
 CC typing, forensic biology), predictive medicine (e.g., diagnostic assays,
 CC prognostic assays, monitoring clinical trials and pharmacogenomic), and
 CC in methods of treatment (e.g., therapeutic and prophylactic). NOVX is
 CC useful as immunogen to produce antibodies immunospecific for NOVX, as
 CC vaccines to screen for potential agonist and antagonist compounds, and
 CC as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide
 CC is useful in gene therapy, to express NOVX, to detect NOVX mRNA
 CC or a genetic lesion in a NOVX gene, and to modulate NOVX activity. The
 CC vector is useful for producing non-human transgenic animals. The antibody
 CC is useful for isolating, and purifying NOVX and to monitor protein levels
 CC in tissue as part of a clinical testing procedure. The present sequence
 CC represents a NOVX protein.

XX Sequence 660 AA:

Query Match 84.6%; Score 898; DB 23; Length 660;
 Best Local Similarity 89.0%; Pred. No. 1.2e-89;
 Matches 178; Conservative 1; Mismatches 5; Indels 16; Gaps 4;

QY 2 SOPANQTSFGATEDEYSXSWYIDEPQGXELQPEGEVPSCHTSTIPGLYHACLASLS-- 59
 DB 3 SOPANQTSFGATEDEYSXSWYIDEPQGEELQPEG-VPSCHTSTIPGLYHACLASLSFQ 61
 QY 60 ILVLLLLMLVRRQLMPDVCVRGRLPSVPVDFLAGDRPRAVPAVFWLLSSLCLLLPD 119
 DB 62 ILVLLLLMLVRRQLMPDCCGHL-----SPVDFLAGDRRAVPAVFWLLSSLCLLLPD 117
 QY 120 EDALPFLTLASPSQDGKTEAPRGAMKILGFHYALYPLAACATAGHTAHLIGSTLS 179
 DB 118 EDALPFLTLASAPS-----GAMKILGFHYALYPLAACATAGHTAHLIGSTLS 168
 QY 180 WAHLGVQVWORAECPOVPKI 199
 DB 169 WAHLGVQVWORAECPOVPKI 188

RESULT 13
 AAB94108
 ID AAB94108 standard; Protein: 560 AA.

AC AAB94108;
 DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:14340.

KW Human; primer: detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

PS Claim 8; SEQ ID 14340; 2537pp + CD ROM, English.

XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides, and the combination of
 CC the 5'-end sequence and 3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 560 AA;

Query Match 46.0%; Score 488; DB 22; Length 560;
 Best Local Similarity 98.9%; Pred. No. 1e-44;

Matches 92; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 107 MYLSSLCCLLPDEBDALPFLTLASAPSQDGKTEAPRGAMKILGFHYALYPLAACATA 166
 DB 1 MYLSSLCCLLPDEBDALPFLTLASAPSQDGKTEAPRGAMKILGFHYALYPLAACATA 60

QY 167 GHTAHLIGSTLSWAHLGVQVWORAECPOVPKI 199
 DB 61 GHTAHLIGSTLSWAHLGVQVWORAECPOVPKI 93

RESULT 14

ID AABG02155 standard; Protein: 1070 AA.

AC AABG02155;
 DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #2146.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS66342.

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity -

PS Claim 20; SEQ ID No 32514; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as

a food supplement. (ii) and its binding partners are useful in medical imaging of sites expressing (ii). (i) and (ii) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIP0 at ftp.wipo.int/pub/published_pct_sequences.

Sequence 1070 AA;

Query Match 8.2%; Score 87.5; DB 22; Length 1070;
Best Local Similarity 25.1%; Pred. No. 2.1;
Matches 50; Conservative 12; Mismatches 70; Indels 67; Gaps 7;

QY 7 NOTSPGATEDYSGSWYID---EPQGXELOPEGEVPSCHTSIPPLVYH--ACLASLSTL 61
DB 633 NUDERPAESDYTSSTSYSEQPNGLPGRHKIPESQPERSTLFGSLPVAAGS----- 687
QY 62 VLLLLAMLVRRQLWPDVCRGRGLSPVDFLAGDRPRA----- 100
DB 688 VLSGSPLEGESSPSHRSRTVSAS---TGDLPKASKIDEPALQTAHSAAPLVPA 743
QY 101 ---VPAVFMVLLSLCLLPDEDLFPVLTLASAPS-----QDQK 137
DB 744 VVNCPTPCSMVLLLECLMLTRSVSPHSPSSSTAVVTVESVHPCTTSTIQEAK 803
QY 138 TEA-----PRGAWKI 147
DB 804 TRAADLVNPLDPRNDKI 822

RESULT 15

AAR48063
ID AAR48063 standard; Protein; 380 AA.

AAR48063;

20-JUL-1994 (first entry)

DE Sequence of polypeptide encoded by the first open reading frame in
DE the unique short (US) region of bovine herpes virus (BHV) genome.

XX Insertion region; unique short region; US; vaccine; antigen.

OS Bovine herpes virus type 1, strain ST.

XX WO9400586-A.

XX 06-JAN-1994.

XX 25-JUN-1993; 93WO-FR00642.

XX 26-JUN-1992; 92FR-0007930.

XX (INNR) RHONE MERIEUX SA.

XX Audonnet JF, Legasseletois ICMA, Leung-tack P, Riviere MEA;

XX WPI; 1994-026222/03.

XX N-PSDB; AA053350.

PT New insertion region sequence of bovine herpes virus genomic DNA
PT - used for recombinant virus with this region deleted or
PT inactivated, useful in vaccines allowing differentiation between
PT vaccinated and infected cattle

XX Claim 9; Fig 2; 47pp; French.

CC The 4190 bp sequence in AA053350 encodes polypeptides homologous to
CC HSV-1 gI, gE and US9; these are neither essential for in vitro
CC replication nor important in inducing a protective immune response.
CC It is genomic DNA purified from the ST strain of BHV-1. (BHV is also
CC known as infectious bovine rhinotracheitis virus). The sequence
CC forms the insertion region in genomic DNA. BHV in which the
CC specific insertion region, esp. nucleotides 172-1311, has been
CC deleted or inactivated by insertion are claimed.

Sequence 380 AA;

Query Match 8.1%; Score 85.5; DB 15; Length 380;
Best Local Similarity 27.6%; Pred. No. 0.95;
Matches 37; Conservative 12; Mismatches 38; Indels 47; Gaps 6;

QY 4 PAGNOTSPGATEDYSGSWYIDEPQGXELOPEGEVPSCHTSIPPLVYHACTAS---LS 59
DB 256 PAGPTASPAASPSRFAFA-----AAPAAAOPADDTTA-----RFRRLASILVPLC 302
QY 60 ILVLLLLAM-----LVRRRL-----WPDVCRGRGLSPV 90
DB 303 VLVLVLLALCAATVNCALRRRLPCSRRVYKPTCAACGSGTCAGRPCCGAAPSAPATV 362
QY 91 DFLAGDRPRAVPAA 104
DB 363 VAL-GSRPKAPPLA 375

Search completed: March 13, 2003, 11:28:52
Job time : 29.0649 secs

copy of 44 up!
% 63 may 00
copy up removed
removed '0008

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:26:32 ; Search time 34.1126 Seconds
(without alignments)
1963.067 Million cell updates/sec

Title: US-09-016-653A-4
Perfect score: 1695
Sequence: 1 CYISALVLSCLLTFLVMRS.....APPTLMESQGFWRRKLVG 325

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
```

```

1: SPtrembl_21:*
2:  sp.archaea:*
3: 2:  sp.bacteria:*
4: 3:  sp.fungi:*
5: 4:  sp.insecta:*
6: 5:  sp.invertebrate:*
7: 6:  sp.mammal:*
8: 7:  sp.mhc:*
9: 8:  sp.ornithella:*
10: 9:  sp.phage:*
11: 10: sp.plant:*
12: 11: sp.podent:*
13: 12: sp.virus:*
14: 13: sp.vertebrate:*
15: 14: sp.unclassified:*
16: 15: sp.virus:*
17: 16: sp.bacteriap:*
18: 17: sp.archaea:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1110.5	65.5	667	4	Q9BX79	O9bx9 homo sapien
2	1110.5	65.5	657	4	Q8RT21	O8rb1 homo sapien
3	1106.5	65.3	658	4	Q9BX78	O9bx8 homo sapien
4	832.5	49.1	670	11	Q7O491	O7O41 mus musculus
5	133	7.8	621	11	Q9DBN1	Q9dbn1 mus musculus
6	98.5	5.8	1841	11	Q9ER60	Q9er60 mus musculus
7	97.5	5.8	403	4	O75205	O75205 homo sapien
8	96.5	5.7	593	16	O83179	O83179 homo sapien
9	92	5.4	796	3	Q9P3B4	Q9p3B4 leprospora
10	92	5.4	2008	4	Q9ES58	Q9es58 homo sapien
11	92	5.4	2146	4	Q9NR73	O9nr73 homo sapien
12	92	5.4	2146	4	Q9BZC4	O9bzC4 homo sapien
13	91.5	5.4	581	1	O07289	O07289 natronomona
14	91	5.4	379	8	Q8WD58	Q8wd58 lepidosiren
15	90.5	5.3	229	8	Q37632	Q37632 thagolectis
16	90.5	5.3	229	8	Q37636	Q37636 thagolectis

17	90.5	5.3	320	8	OBMA20	OBMA20	ophiobagrus
18	90.5	5.3	400	2	OBVNT4	OBVNT4	enterobacter
19	90.5	5.3	730	12	Q06634	Q06634	bovine herp
20	90	5.3	418	4	Q8TDR0	Q8TDR0	homo sapien
21	90	5.3	474	4	Q9BUE5	Q9BUE5	homo sapien
22	90	5.3	476	4	Q9BXP3	Q9BXP3	homo sapien
23	89.5	5.3	107	16	OBXY12	OBXY12	escherichia
24	89.5	5.3	318	8	OBMA25	OBMA25	matricora bl
25	89.5	5.3	319	8	OBMA18	OBMA18	pseudaspis
26	89.5	5.3	361	16	OBXTH6	OBXTH6	raletonia s
27	89	5.3	1082	10	Q8RWU0	Q8RWU0	sorghum bicc
28	89	5.3	3164	12	Q69088	Q69088	human herpe
29	88.5	5.2	374	8	Q21210	Q21210	parachromis
30	88.5	5.2	395	16	Q8Z119	Q8Z119	salmonella
31	88.5	5.2	402	16	Q8ZMH0	Q8ZMH0	salmonella
32	87.5	5.2	226	8	Q20419	Q20419	hodotermops
33	87.5	5.2	229	8	Q9ZXX2	Q9ZXX2	panulirus a
34	87.5	5.2	229	8	Q37663	Q37663	rhagoletis
35	87.5	5.2	229	8	Q37439	Q37439	carpomya sh
36	87.5	5.2	318	8	OBMA21	OBMA21	naja nivea
37	87.5	5.2	367	4	Q96MS7	Q96MS7	homo sapien
38	87.5	5.2	473	8	Q21104	Q21104	archocentru
39	87.5	5.2	463	10	Q9AS50	Q9AS50	oryza sativ
40	87.5	5.2	554	6	Q95L59	Q95L59	bos taurus
41	87.5	5.2	580	6	Q8MW14	Q8MW14	bos taurus
42	87	5.1	2015	4	Q96J69	Q96J69	homo sapien
43	87	5.1	2159	11	Q91V24	Q91V24	mus musculi
44	86.5	5.1	2454	2	Q05135	Q05135	pseudomonas
45	86.5	5.1	301	16	Q915M4	Q915M4	pseudomonas

ALIGNMENTS

RESULT 1

ID	Q9BX79	PRELIMINARY;	PRT;	667	AA
----	--------	--------------	------	-----	----

Query Match 65.5%; Score 1110.5; DB 4; Length 667;

DT 01-JUN-2001 (TREMBLrel_17, Created)
DT 01-JUN-2001 (TREMBLrel_17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel_21, Last annotation update)
DE STRA6 isoform 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Szeto W., Jiang W., Tice D.A., Rubinfeld B., Hollingshead P.G.,
RA Fong S.E., Dugger D.L., Plam T., Yansura D.G., Wong T.A.,
RA Grimaldi J.C., Corpuz R.T., Singh U.S., Frantz G.D., Devaux B.,
RA Crowley C.W., Schwall R.H., Eberhard D.A., Raetzel L., Polakis P.,
RA Pennic D.;
RT "Overexpression of the retinoid acid-responsive gene Stra6 in human
RT cancers and its synergistic induction by Wnt-1 and retinoic acid.";
RL Cancer Res. 0:0-0(2001).
LN [2]
RN SEQUENCE OF 108-667 FROM N.A.
RP Isegaki T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Masatsuka T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Matsutsuma M., Hosogli T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takehashi M., Chiba K., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Wakabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Maeno Y.,
RA Nimomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF532728; AAK30289.1;-
DR EMBL; AK022603; BAB14122.1;-
FT VARIANT 527 M -> I.
SQ SEQUENCE 667 AA; 73502 MW; D20840A6998BA2E CRC64;

Best Local Similarity 76.6%; Pred. No. 4,9e-90;
Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYISALVLSCLTFLVLMRSVLTNR-----LGSQSGSD-----GQFSWMLFS 42
DB 365 CYISALVLSCLTFLVLMRSVLTNR-----LGSQSGSD-----GQFSWMLFS 424
QY 43 VPLPLPPLAGLLVVOOIIFFLGTTLAFLVLMVPLHGRNLLFRSLSSSPFWLTALAVI 102
DB 425 AYOTAFICGLIVVOOIIFFLGTTLAFLVLMVPLHGRNLLFRSLSSSPFWLTALAVI 484
QY 103 LONMAAHVFLETHDGHQPOLTNRRVLYAATFLLPPLNVLVGAMXXXC----- 149
DB 485 LONMAAHVFLETHDGHQPOLTNRRVLYAATFLLPPLNVLVGAMXXXC----- 149
QY 150 -SPSIAIRHP-----TPGYTYRNFLEKIEVSQSHPAMTAFCSLLQAOSSLPRMAAPD 203
DB 545 GOMDSLPRPRAATLDPGYTYRNFLEKIEVSQSHPAMTAFCSLLQAOSSLPRMAAPD 604
QY 204 SLRPEDEGMOQLQOTKDSMAKGPAGXRGARWGLAYTLLHNPTLOVFRKTALLGANG 263
DB 605 SLRPEDEGMOQLQOTKDSMAKGPAGXRGARWGLAYTLLHNPTLOVFRKTALLGANG 664
QY 264 AQP 266
DB 665 AQP 667

RESULT 2

Q8TB21 PRELIMINARY; PRT; 667 AA.

AC 08TB21;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 73.5 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Straube R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025256; AA825256.1; -
KM Hypothetical protein.
SQ SEQUENCE 667 AA; 73532 MW; 120665F4635CB0E5 CRC64;

Query Match 65.5%; Score 1110.5; DB 4; Length 667;
Best Local Similarity 76.6%; Pred. No. 4,9e-90;
Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYISALVLSCLTFLVLMRSVLTNR-----LGSQSGSD-----GQFSWMLFS 42
DB 365 CYISALVLSCLTFLVLMRSVLTNR-----LGSQSGSD-----GQFSWMLFS 424
QY 43 VPLPLPPLAGLLVVOOIIFFLGTTLAFLVLMVPLHGRNLLFRSLSSSPFWLTALAVI 102
DB 425 AYOTAFICGLIVVOOIIFFLGTTLAFLVLMVPLHGRNLLFRSLSSSPFWLTALAVI 484
QY 103 LONMAAHVFLETHDGHQPOLTNRRVLYAATFLLPPLNVLVGAMXXXC----- 149
DB 485 LONMAAHVFLETHDGHQPOLTNRRVLYAATFLLPPLNVLVGAMXXXC----- 149
QY 150 -SPSIAIRHP-----TPGYTYRNFLEKIEVSQSHPAMTAFCSLLQAOSSLPRMAAPD 203
DB 545 GOMDSLPRPRAATLDPGYTYRNFLEKIEVSQSHPAMTAFCSLLQAOSSLPRMAAPD 604
QY 204 SLRPEDEGMOQLQOTKDSMAKGPAGXRGARWGLAYTLLHNPTLOVFRKTALLGANG 263
DB 605 SLRPEDEGMOQLQOTKDSMAKGPAGXRGARWGLAYTLLHNPTLOVFRKTALLGANG 664

QY 264 AQP 266
DB 665 AQP 667

RESULT 3

Q9BX78 PRELIMINARY; PRT; 658 AA.

AC Q9BX78;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE STRA6 isoform 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Szeto W., Jiang W., Tice D.A., Rubinfeld B., Hollingshead P.G.,
RA Fong S.E., Dugger D.L., Pham F., Yansura D.G., Wong T.A.,
RA Grimaldi J.C., Corpuz R.T., Singh J.S., Frantz G.D., Devaux B.,
RA Crowley C.W., Schwall R.H., Eberhard D.A., Raetzell L., Polakis P.,
RA Pennica D.;
RT "Overexpression of the retinoic acid-responsive gene Stra6 in human
RT cancers and its synergistic induction by mnt-1 and retinoic acid.";
RL Cancer Res. 0:0-0(2001).
DR EMBL; AF352729; AA830290.1; -
FT VARIANT 518 518 I -> M.
SQ SEQUENCE 658 AA; 72582 MW; 5405281727F647DD CRC64;

Query Match 65.3%; Score 1106.5; DB 4; Length 658;
Best Local Similarity 76.2%; Pred. No. 1,1e-89;
Matches 231; Conservative 5; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYISALVLSCLTFLVLMRSVLTNR-----LGSQSGSD-----GQFSWMLFS 42
DB 356 CYISALVLSCLTFLVLMRSVLTNR-----LGSQSGSD-----GQFSWMLFS 415
QY 43 VPLPLPPLAGLLVVOOIIFFLGTTLAFLVLMVPLHGRNLLFRSLSSSPFWLTALAVI 102
DB 416 AYOTAFICGLIVVOOIIFFLGTTLAFLVLMVPLHGRNLLFRSLSSSPFWLTALAVI 475
QY 103 LONMAAHVFLETHDGHQPOLTNRRVLYAATFLLPPLNVLVGAMXXXC----- 149
DB 476 LONMAAHVFLETHDGHQPOLTNRRVLYAATFLLPPLNVLVGAMXXXC----- 149
QY 150 -SPSIAIRHP-----TPGYTYRNFLEKIEVSQSHPAMTAFCSLLQAOSSLPRMAAPD 203
DB 536 GOMDSLPRPRAATLDPGYTYRNFLEKIEVSQSHPAMTAFCSLLQAOSSLPRMAAPD 595
QY 204 SLRPEDEGMOQLQOTKDSMAKGPAGXRGARWGLAYTLLHNPTLOVFRKTALLGANG 263
DB 596 SLRPEDEGMOQLQOTKDSMAKGPAGXRGARWGLAYTLLHNPTLOVFRKTALLGANG 655
QY 264 AQP 266
DB 656 AQP 658

RESULT 4

Q70491 PRELIMINARY; PRT; 670 AA.

AC Q70491;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE Retinoic acid-responsive protein.
GN STRA6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95377533; PubMed=7649373;
 RA Bouillet P., Oulad-Abdelghani M., Vicaire S., Garnier J.M.,
 RA Schaubert B., Dolle P., Chambon P.;
 RT "Efficient cloning of cDNAs of retinotic acid-responsive genes in P19
 RT embryonal carcinoma cells and characterization of a novel mouse gene,
 RT Stral (mouse LERK-2/Bp1g2)."
 RL Dev. Biol. 170:420-433(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97346723; PubMed=9203140;
 RA Bouillet P., Sapin V., Chazaud C., Mesaddeq N., Decimo D., Dolle P.,
 RA Chambon P.;
 RT "Developmental expression pattern of Stra6, a retinotic acid-responsive
 RT gene encoding a new type of membrane protein."
 RL Mech. Dev. 63:173-186(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Bouillet P.;
 RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF062476; AAC16016.1; -.
 DR MGI; MGI:107742; Stra6.
 SQ SEQUENCE 670 AA; 73774 MW; 6ACE69F3EE16A4F7 CRC64;

 Query Match 49.1%; Score 832.5; DB 11; Length 670;
 Best Local Similarity 57.7%; Pred. No. 2e-65; Mismatches 62; Indels 39; Gaps 4;
 Matches 176; Conservative 28; Mismatches 62; Indels 39; Gaps 4;

 QY 1 CYIALVLSCLTFLVLRSLVTHR-----LGSQSGSD-----GQFSWNLFS 42
 DB 366 CYIALVLSCLTFLVLRSLVTHRSLRTHRANLQALHGAALDLPLOSHPRSQAVWSMFC 425
 QY 43 VPLPLPLAGLIVQIIFFLGTTALAFVLMPLVHGRNLLFRSLSSNPFLTALAVI 102
 DB 426 AVQAFSCGLVQVIFFLGTTSLAFVFLVPLHGRNLLFRSLSSNPFLTALAVI 485
 QY 103 LONMAAHVFLTHGHQDPLTRRVLYATFLPLVNVVGAMXXXCSPSIAIRPT-- 159
 DB 486 LONIAAHVFLTHGHQDPLTRRVLYATFLPLVNVVGAMXXXCSPSIAIRPT-- 159
 QY 160 -----PGYTYRNFLEKIEVSQSHPMATFCSLLQAQSLPRTMAAPD 203
 DB 546 GQMDLSLPLQRAASLDPGYHTYQNFIRIASQSHPCVIAFCALLHAPSPQRPPLAPD 605
 QY 204 SLRPGSEDEGMQLQTKDSMAKGPARGKARWGLATTLHNPLOVFRRTALLG--A 261
 DB 606 SLRPAEEBEGMQLQTKDIMAKGAGHKGSQSRARWGLATTLHNPLOVFRRTALLG 665
 QY 262 NGAQP 266
 DB 666 NGTOP 670

 RESULT 5
 Q9DBN1 PRELIMINARY; PRT; 621 AA.
 AC Q9DBN1;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE 1300002K09R1K.
 GN 1300002K09R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=LIVER;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh J., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Glass C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsumoto Y., Nikaido I., Penzo G., Quackenbush J.,
 RA Schmitt L.M., Staudt P., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baladaoui R., Barh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gasterich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Sakamoto N.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohdeki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK004855; BAB23619.1; -.
 DR MGI; MGI:1921402; 1300002K09R1K.
 SQ SEQUENCE 621 AA; 71002 MW; 909E1C6BC2A9EA60 CRC64;

 Query Match 7.8%; Score 133; DB 11; Length 621;
 Best Local Similarity 23.1%; Pred. No. 0.0015;
 Matches 55; Conservative 37; Mismatches 92; Indels 54; Gaps 7;

 QY 50 LAGLIVQIIFFLGTTALAFVLMPLVHGRNLLFRSLSSNPFLTALAV--ILQNM 106
 DB 388 LMGTLIHVVQSLQGVLMVGLVPIIHRGL--EMLQGFGLVGLTIVGLILQVW 444
 QY 107 AAHVFLTHGHQD-----LNNRVLYATFLPLVNVVGAMXXXCSPSIAI----- 155
 DB 445 IAGFFLQPKLGTSDXKPLANNRRAFNFFLFYNNVLLD--LQACLSRLISCLLG 502
 QY 156 -----RHPTGYTYRNFLEKIEVSQSHPMATFCSLLQAQSLPRTMA 199
 DB 503 TWLIARIDRTIMOSGVGAMGFGAMIGMLFVDHYHNPLVVFCHILLIS----- 553
 QY 200 APQDSLRPGSEDEGMQLQTKDSMAKGPARGKARWGLATTLHNPLOVFRRTALLG 256
 DB 554 -----HKDRLQGTQVYVYVCLNQAGRFARTRFWFLQTLINBRLVLRKS 602

 RESULT 6
 Q9ER60 PRELIMINARY; PRT; 1841 AA.
 AC Q9ER60;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Voltage-gated sodium channel.
 GN SCN4A.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA1B/C; TISSUE=HEART;
 RA Zimmer T., Bendorf K.;
 RT "The mouse heart sodium channels: cloning and characterization of two
 RT distinct isoforms and alternatively spliced variants."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ278787; CAC17146.1; -.
 DR MGI; MGI:98250; Scn4a.
 DR InterPro; IPR001682; Cat/Na pore.
 DR InterPro; IPR002111; Cat_channel_Tyrl.
 DR InterPro; IPR000048; IQ_region.
 DR InterPro; IPR000636; M+channel_nlg.
 DR InterPro; IPR001915; PKD 2.
 DR Pfam; PF00520; Ion trans_4.
 DR Pfam; PF00612; IQ_1.

DR PRINTS; PRO0170; NACHANNEL.
 DR PRINTS; PRO1433; POLYCYSTRIN2.
 DR SMART; SM00015; IQ; 1.
 KW Ionic channel.
 SQ SEQUENCE 1841 AA; 208797 MW; 0766PDD33A9E0E55 CRC64;
 Query Match 5.8%; Score 98.5; DB 11; Length 1841;
 Best Local Similarity 22.0%; Pred. No. 5.4;
 Matches 67; Conservative 35; Mismatches 105; Indels 97; Gaps 14;

QY 35 QFWMNLFVSLPLPPLAGLLVQOIIFFLGTALAFVLMVHLGRLNLFRRSLSSWPMW 94
 DB 633 QOCMNIFD-----SFIVTLSTVELG---LANVQGLSVLSFRLVFLAKSWP-T 679
 QY 95 LTLALAVIIONMAAHWFLETHDGHQPLTRRYLYAATFLFPL--NLVGMAMXXXCSS 152
 DB 680 LNNLIIKINGSVCA-----LSNLTIVLAIIFIFVAVGMOLPGMSYKCVCK 726
 QY 153 IAIKHPPTGYT---YRNPLKI-----EVSQSHPMATFCSLLLOAQ-- 191
 DB 727 IASDCSLPRMHMDFHSLFVRLICGEWETMDQMEVAGQAMCLTFLVMVYIGNLV 786
 QY 192 -----SLPRTMAAPQDSLRPEDEGQULQTKDSMAKGAAPKXGRARWGLAYT-- 243
 DB 787 VNLFLALILSSFA--DSLAADEDEGEMNNLQ-----IAIGRIKMGIAFAKT 832
 QY 244 ----LHNPTLOVFRKTLT-----GANGAQPCCSLPGSPSITPAMQAPPDHXG 291
 DB 833 FLGLGHKILSL--KDIWLSLGEPCGAGENGESP-----PDEKKEPPEDG 878
 QY 292 XVEV 295
 DB 879 NKEI 882

RESULT 7
 075205 PRELIMINARY; PRT; 403 AA.
 AC 075205; QGNZHO;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 44.8 kDa protein (GPCR5B protein) (Orphan G-protein coupled receptor).
 GN A-69G12.1 OR GPCR5B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Loftus B.J., Zhou L., Crosby M., Fuhrmann J., Mason T.M., Brandon R., Kim U.J., Kexlavage A.R., Venter J.C.;
 RT "Homo sapiens Chromosome 16 BAC clone C1787SK-A-69G12";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Wei H., Osborne B., Spruyt M., Murphy D.;
 RT "Cloning of a novel G protein-coupled receptor localized on human chromosome 16p12.";
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Robbins M.J., Michalovich D., Hill J., Calver A.R., Medhurst A.D., Gloger I., Sims M.A., Middlemiss D.N., Pangalos M.N.;
 RT "Molecular cloning and characterisation of two novel retinoic acid inducible orphan G-protein coupled receptors.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 9-403 FROM N.A.
 RX MEDLINE=20247251; PubMed=10783259;
 RA Brauner-Oborne H., Krosgaard-Larsen P.;

RT "Sequence and expression pattern of a novel human orphan G-protein-coupled receptor, GPCR5B, a family C receptor with a short amino-terminal domain.";
 RT Genomics 65:121-128 (2000).
 RL EMBL; AC004131; AAC7544.1; -
 DR EMBL; AF181862; AAF0531.1; -
 DR EMBL; AJ276101; CAC00632.1; -
 DR EMBL; AF202640; AAF67321.1; -
 DR InterPro; IPR002956; Bridge of 7less.
 DR InterPro; IPR000337; GPCR_Mgr.
 DR Pfam; PF00003; 7tm 3; 1.
 DR PRINTS; PRO1223; BRIDGEOFLSS.
 DR PROSITE; PS50253; G_PROTEIN_RECEP_F3_4; 1.
 KW Hypothetical protein; Receptor.
 SQ SEQUENCE 403 AA; 44795 MW; 3902A16CA69C36E CRC64;
 Query Match 5.8%; Score 97.5; DB 4; Length 403;
 Best Local Similarity 19.9%; Pred. No. 1.3; Mismatches 107; Indels 141; Gaps 14;
 Matches 69; Conservative 30;

QY 5 ALVLSCLTFLVLMRSLVTRHLSGSGSDQFSWNLFSVPLPLPPLAGLLVQOII----- 59
 DB 134 ALCFSCILSQAMRVRLVRH-----GTGPAGQLVGLALCL-----MLVQVIAVBEWL 181
 QY 60 -----FLLGTALAFVLMR-----VLHGRNLFRRS 86
 DB 182 VLTVLRTPRACAYEPMDFVMAIYDNLVLTITGLTLFLTCGFKWKILNGAFLLITAF 241
 QY 87 LE-SWPFWLT-----ALAVIIONMAAHWFLETHDGHQPLTRNRYL 128
 DB 242 LSVLIWVAWMTMYLFGNVKLQGGDAMNDPLTALITL--AASGWFIYFH-AIPEI----- 232
 QY 129 YAAFTFLPLNLVGMAMXXXCSPSIAIRHPTPGYTYTRNLFKIEVSQSHPMATFCSLL 188
 DB 293 -----HCTLLPALQENTPVYF-----DTSQPRMRETAFA----- 320
 QY 189 QAQSLPRTMAAPQDSLRPEDEGQULQTKDSMAKGAAPKXGRARWGLAYTLLHP 248
 DB 321 EEDVQLPRAY-----ENKAFSMDENANLRTAGCPNGSLCK----- 357
 QY 249 TLQVFRKTLALGANGAQPCCSLPGSPSITPAMQ-----PAGPPDHXG 291
 DB 358 -----RPSGSLGKRPSAFRFSNVYQPTFEMAVLVNGGITPAPPSHNG 399

RESULT 8
 083179 PRELIMINARY; PRT; 593 AA.
 AC 083179;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Thiamine ABC transporter, permease protein, putative.
 GN TP0143.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 OC NCBI_Taxid=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NICHOLS;
 RX MEDLINE=98332770; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G., Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A., Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J., Khakhria H., Richardson D., Howell J.K., Childsbaram M., Ditterback T., McDonald L., Attalach P., Bowman C., Cotton M.D., Fujii C., Garland S., Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O., Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis spirochete.";
 RL Science 281:375-388 (1998).
 DR EMBL; AB001199; AAC65131.1; -
 DR TIGR; TP0143; -.

DR InterPro: IPR000515; BPD_transp.
DR Pfam: PF00528; BPD_transp. 1.
KW Complete proteome.
SQ SEQUENCE 593 AA; 64882 MW; 5D2E3F9189A9FCB CRC64;

Query Match 5.7%; Score 96.5; DB 16; Length 593;
Best Local Similarity 23.0%; Pred. No. 2.4;
Matches 61; Conservative 29; Mismatches 86; Indels 89; Gaps 12;

QY 3 ISALVLSCLT---TFLVLRSLVTHRLSGSGSGDGFNNLPSVPLPPLAGLVQOI 58
DB 120 LCSAVLACAVGFAAFVFKSLARALFPA-----LCAIPAIPL----- 161
QY 59 IFFIGTALAPLVIMPLHGRNLLFFRSLSSEWPF-----WLTALA 100
DB 162 -----TVALAFLLSF-----GNGGLCTRTLTGLTGVSTPRTFVLSASGVIAHAMYNPPLA 212
QY 101 VILQNMMAHWFLETHDGHPOLTNR--RVLYAATFLLFPLNVLGAMXXX-----CSP 151
DB 213 LAIARAWNTISADAEDARLLGARAVRYFTIT-----LPALAGALRSSFLVIFLYCFP 267
QY 152 SIAT-----RHPTPGYTYRNFLEKIEVSQHP-----AMTAFCSLLI----- 188
DB 268 SLMMVLLIGTFTTLEVELYRS---IRTOAHHPYASTLALSTFYALLMIMGVSTESQ 324
QY 189 QAQSILPRTMAAPQDSLRGDEDEG 213
DB 325 QAQSCVARTRPLPQSIQGNIERIG 349

RESULT 9

ID Q9P3B4 PRELIMINARY; PRT; 796 AA.
AC Q9P3B4;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Related to ser/thr protein kinase IME2.
GN B7N4.100.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariiales; Sordariaceae; Neurospora.
CX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hohelsel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.,
RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RA Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AL390218; CAB99241.2; -.
DR HSSP; P24941; 1BUH.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002280; Ser_thr_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRKC; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; UNKNOWN_1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 2.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 796 AA; 87388 MW; 69B1C17DEBB97AB CRC64;

Query Match 5.4%; Score 92; DB 3; Length 796;
Best Local Similarity 20.7%; Pred. No. 8.3;
Matches 75; Conservative 38; Mismatches 131; Indels 118; Gaps 14;
QY 23 THRLSGSGS-----GDGQFSWNLPSVPLPPLAGLVQO--IIFLGTTALAP 69

DB 13 SHGIGSGQALEDRFEVLKEIDGSG-----SVVLARVSAGATVARRGCTVIAITMKTF 68
QY 70 LVLMPLVHGRNLLFFRSLS-----WPFVLTALAVILQNMMAHWFLETHD 118
DB 69 ESVGQCMELREVVFLRTLPAPHVLPALDIFLDPF--TKKHIMEVMEGNLYQLMKARD 126
QY 119 HPQLTNRVY-----LVYATFL--LPELVNVLGAMXXXCSPSIAIHN----- 157
DB 127 HKCLDNSSVKSILFQIMKGLEHIAHHEFRDIPKENILVSTSSHMDATNFRFRSALMN 186
QY 158 --PTPGYTYRNFLEKIEVSQHPATAFCSLLQAOSSLP-----RTMAPODSLRG 208
DB 187 PPPTPTTYVX-----IADFGARETHSKLPPTTYVSTWYRAPEVLPAG 232
QY 209 EDEEGMQL-----LOTQSMAGARPGAXGRARWGLAYTLHINFTLOVFRKALLGA-- 261
DB 233 EYSAPVDIAIGMAVEIATIKLPFG-----NEVDQVMRYCEIMSGSPG 277
QY 262 -----NGAQPCSSLGSPSITP-----AMQAPGPDHGXVEVCLHW 299
DB 278 NMYNKAGARVGGWREGTRLAGKLGFSFPKMAPHSMDTIIQTPQWPASLAHFTVWCLMW 337
QY 300 EP 301
DB 338 DP 339

RESULT 10

ID Q96S58 PRELIMINARY; PRT; 2008 AA.
AC Q96S58;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE ABCA-SSN.
GN ABCA7/ABCA-SSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2125283; PubMed=11355874; Arakawa R., Sadanami K.,
RA Tanaka A., Ikeda Y., Abe-Dohmae S.,
RA Kidera A., Nakagawa S., Nagase T., Aoki R., Kioka N., Amachi T.,
RA Yokoyama S., Ueda K.;
RT "Human ABCA1 Contains a Large Amino-Terminal Extracellular Domain
RT Homologous to an Epitope of Sjogren's Syndrome";
RL Biochem. Biophys. Res. Commun. 283:1019-1025(2001).
DR EMBL; AB055390; BAB62294.1; -.
DR InterPro: IPR003439; ABC_transport.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transport; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 2008 AA; 218617 MW; 226F85C24230B90 CRC64;

Query Match 5.4%; Score 92; DB 4; Length 2008;
Best Local Similarity 25.3%; Pred. No. 22;
Matches 80; Conservative 48; Mismatches 126; Indels 62; Gaps 17;

QY 3 ISALVLSCLT---LTF-----LVLRSLVT---HRLSGSGSD-----GQFSNVL--FSV 43
DB 1396 VDVAVSICVVPAMSFVPASFTLVLIBERVTRAKLIQLMGGLSPTLYWLGNFLLMCMNYLV 1455
QY 44 PLPLPPLAGLVQOIIIFLGTTALAPLVIMPLHGRNLL-----FFRSLSSEWPFMLT 96
DB 1456 PACIVVLLFLFPQQA--VAPANLPALDLVLVGMSTPTPLMYPASFVSPSTAYVVL 1514
QY 97 -LALAVILQNMMAHWFLETHDGHPOLTNRRVLYAATFLLFPLNVL--VGAMXXXCSPSIA 154

Db 1515 CINLFIGINGSMATFVLEFSDOKLOEVS-ILKQVFLIFPHFCIGRLIDMVRNOA 1572
 QY 155 IRHPTGTYRNFLKIEVSOSH-----PAMTAFCSLLLOAOSL--PRTMAAPD 203
 Db 1573 DAFERLGDROFQSPFLRWEVGNKLLAMVIOGPLFLF-TLLQHSQSLPQPRVSLP- 1629
 QY 204 SLRGEDEGMOQLQTKDSMAKGARPG-----AKRGARMGCLATYLLHNPTLOVF 253
 Db 1630 --LLGEDE--DVARERERVQATOGDVLVLRNLTQYVRGQMPAVDRLCIGIPGECF 1685
 QY 254 RKTALLGANGAOPCSS 269
 Db 1686 ---GLLGAVNGAGKTS 1698

RESULT 11
 Q9NR73
 ID Q9NR73 PRELIMINARY; PRT; 2146 AA.
 AC Q9NR73; 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Macrophage ABC transporter.
 GN ABCA7.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20334305; PubMed=10873640.
 RA Kaminski W.E., Orvo E., Diederich W., Klucken J., Drobnik W.,
 Schmitz G.,
 RT "Identification of a Novel Human Sterol-Sensitive ATP-Binding Cassette
 Transporter (ABCA7)."
 RL Biochem. Biophys. Res. Commun. 273:532-538(2000).
 DR EMBL; AF250238; AAF85794.1;
 DR InterPro; IPR003593; AAA_Atpase.
 DR InterPro; IPR003439; ABC_transport.
 DR Pfam; PF00005; ABC_tran; 2.
 DR ProDom; PD000006; ABC_transport; 2.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW ATP-binding.
 SQ SEQUENCE 2146 AA; 234469 MW; 679816EB2D75FF0D CRC64;

Query Match 5.4%; Score 92; DB 4; Length 2146;
 Best Local Similarity 25.3%; Pred. No. 24;
 Matches 80; Conservative 48; Mismatches 126; Indels 62; Gaps 17;

QY 3 ISALVLSCL---LTF-----LVLMRSVLT--HRLGSGSGSD-----GQFSNUL--FSV 43
 Db 1534 VDVLVSLCVFAMSPVPSFTLVLEERTYRAKHLQLMGSLPTLYMGNFLMDMGNVLY 1593
 QY 44 PLPLPPLAGLIVQOIIIFLGTALAFVLMVPLHGRNLL-----FPRSLESSWPFMLT 96
 Db 1594 PACIVVLIFLAFQORA--YVAPANLPLALLLLLLYGWSITPLMPYASFFSVSTAYVLT 1652
 QY 97 -LALAVILQNMAMHVFLETHDGHQPLTNRRVLYATFLFPLNVL--VGAMXXXCSPSIA 154
 Db 1653 CINLFIGINGSMATFVLEFSDOKLOEVS-ILKQVFLIFPHFCIGRLIDMVRNOA 1710
 QY 155 IRHPTGTYRNFLKIEVSOSH-----PAMTAFCSLLLOAOSL--PRTMAAPD 203
 Db 1711 DAFERLGDROFQSPFLRWEVGNKLLAMVIOGPLFLF-TLLQHSQSLPQPRVSLP- 1767
 QY 204 SLRGEDEGMOQLQTKDSMAKGARPG-----AKRGARMGCLATYLLHNPTLOVF 253
 Db 1768 --LLGEDE--DVARERERVQATOGDVLVLRNLTQYVRGQMPAVDRLCIGIPGECF 1823
 QY 254 RKTALLGANGAOPCSS 269

Db 1824 ---GLLGAVNGAGKTS 1836

RESULT 12
 Q9BZC4
 ID Q9BZC4 PRELIMINARY; PRT; 2146 AA.
 AC Q9BZC4; 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE ABC transporter member 7.
 GN ABCA7.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21328888; PubMed=11435699;
 RA Broccardo C., Oborio U., Luciani M.-F., Schriml L.M., Prades C.,
 RA Shulenin S., Arnould I., Naudin L., Lafaygue C., Rosier M., Jordan B.,
 RA Mattei M.G., Dean M., Deneffe P., Chimi G.,
 RT "Comparative analysis of the promoter structure and genomic
 organization of the human and mouse ABCA7 gene encoding a novel ABCA
 transporter."
 RT Cytogenet. Cell Genet. 92:264-270(2001).
 RL EMBL; AF328787; AAK00959.1;
 DR InterPro; IPR003593; AAA_Atpase.
 DR InterPro; IPR003439; ABC_transport.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR Pfam; PF00005; ABC_tran; 2.
 DR ProDom; PD000006; ABC_transport; 2.
 DR SMART; SM00382; AAA; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW ATP-binding.
 SQ SEQUENCE 2146 AA; 234306 MW; 2391728D5AD97E75 CRC64;

Query Match 5.4%; Score 92; DB 4; Length 2146;
 Best Local Similarity 25.3%; Pred. No. 24;
 Matches 80; Conservative 48; Mismatches 126; Indels 62; Gaps 17;

QY 3 ISALVLSCL---LTF-----LVLMRSVLT--HRLGSGSGSD-----GQFSNUL--FSV 43
 Db 1534 VDVLVSLCVFAMSPVPSFTLVLEERTYRAKHLQLMGSLPTLYMGNFLMDMGNVLY 1593
 QY 44 PLPLPPLAGLIVQOIIIFLGTALAFVLMVPLHGRNLL-----FPRSLESSWPFMLT 96
 Db 1594 PACIVVLIFLAFQORA--YVAPANLPLALLLLLLYGWSITPLMPYASFFSVSTAYVLT 1652
 QY 97 -LALAVILQNMAMHVFLETHDGHQPLTNRRVLYATFLFPLNVL--VGAMXXXCSPSIA 154
 Db 1653 CINLFIGINGSMATFVLEFSDOKLOEVS-ILKQVFLIFPHFCIGRLIDMVRNOA 1710
 QY 155 IRHPTGTYRNFLKIEVSOSH-----PAMTAFCSLLLOAOSL--PRTMAAPD 203
 Db 1711 DAFERLGDROFQSPFLRWEVGNKLLAMVIOGPLFLF-TLLQHSQSLPQPRVSLP- 1767
 QY 204 SLRGEDEGMOQLQTKDSMAKGARPG-----AKRGARMGCLATYLLHNPTLOVF 253
 Db 1768 --LLGEDE--DVARERERVQATOGDVLVLRNLTQYVRGQMPAVDRLCIGIPGECF 1823
 QY 254 RKTALLGANGAOPCSS 269
 Db 1824 ---GLLGAVNGAGKTS 1836

RESULT 13
 O07289
 ID O07289 PRELIMINARY; PRT; 581 AA.
 AC O07289;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)

DT 01-JUL-1997 (TReMBLrel. 04, last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
 DE Subunit I-homologue of cytochrome oxidase.
 GN CBA.
 OS Natronomonas pharaonis (Natronobacterium pharaonis).
 OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
 OC Halobacteriaceae; Natronomonas.
 OX NCBI_TaxID=2257;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPI/28;
 RX MEDLINE=98088959; PubMed=9428682;
 RA Møller S., Engelhard M.;
 RT "Cytochrome b3 from Natronobacterium pharaonis: An archaeal four-
 subunit cytochrome-c-type oxidase.";
 RL Eur. J. Biochem. 250:332-341(1997).
 CC -1- COPFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 DR EMBL; Y10500; CAA71525.1; -
 DR HSSP; Q56408; 1BK.
 DR InterPro; IPR000883; COX1.
 DR Pfam; PF00115; COX1.1.
 DR Copper; Heme; Oxidoreductase; Respiratory chain; Transmembrane.
 KW SEQUENCE 581 AA; 63990 MW; 46963269A92D0522 CRC64;
 SQ

Query Match 5.4%; Score 91.5; DB 1; Length 581;
 Best Local Similarity 26.6%; Pred. No. 6.6; Matches 63; Conservative 32; Mismatches 89; Indels 53; Gaps 15;

QY 6 LVLSCLTFLVIMRSVLRHSGSGSDGDFSMVLPPLPPLAGLVQOIIFFLGTT 65
 DB 300 LILSLTLTAFVHVASMEHGARGSGS--GYFGW-LRALPMWDPVVTGALGLMF---A 352
 QY 66 ALALVLMPTVHGNNLFFRSLSWSW---PPMLTLALVILQNNAAWVLETHDHPQ 121
 DB 353 AAARSGM--VNAAGNNINLT--VHNTWVVGHFHLVTGTAVALTFMAVSWYFL-----PQ 402
 QY 122 LTNR-----LVYATFLF-PLNVLVGAMKXCSPIAHPFPGYTYTNPLKIEVSQ 175
 DB 403 ITGKKMGKSYALAAQVLMVFMGTMSNMRSGLAGPRRTAE--QYRNF-EFEMAA 458
 QY 176 SH---PAMTAFCSLLQAQSL-----PRTMAAPQ--DSLPGED 211
 DB 459 GSIGELNAAQVVLGILFLVSTLLFLVYVVMVYLGDKAPGTLPLANEVADTL-SGED 514

RESULT 14

Q8WDS8 PRELIMINARY; PRT; 379 AA.
 AC Q8WDS8;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
 DE Cytochrome b.
 GN CYTB.
 OS Lepidostiren paradoxus (South American lungfish).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
 OC Dipnoi; Lepidostireniiformes; Lepidostirenidae; Lepidostiren.
 OX NCBI_TaxID=7883;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Denk A.G., Binkmann H., Zitzler J., Joes J.J., Meyer A.;
 RT "Complete Mitochondrial Genome Sequences of South American
 RT (Lepidostiren paradoxus) and Australian (Neoceratodus forsteri)
 RT lungfishes: Sarcopneustian - Actinopterygian division confirmed.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF302934; AAL55248.1; -
 DR InterPro; IPR000179; Cyt_b_b6.
 DR Pfam; PF00032; cytochrome_b_C; 1.
 DR Pfam; PF00033; cytochrome_b_N; 1.

DR PROSITE; PS00192; CYTOCHROME_B_HEME; UNKNOWN_1.
 DR PROSITE; PS00193; CYTOCHROME_B_OO; UNKNOWN_1.
 KW Mitochondrion.
 SQ SEQUENCE 379 AA; 42889 MW; 02118DA863322CCB CRC64;

Query Match 5.4%; Score 91; DB 8; Length 379;
 Best Local Similarity 28.1%; Pred. No. 4.6;
 Matches 45; Conservative 20; Mismatches 51; Indels 44; Gaps 9;

QY 11 LILFLVIMRSVLRHSGSGSDGDFSMVLPPLPPLAGLVQOIIFFLGTT 62
 DB 230 LMGFTILMLLIISL-----FSPNLSDPENTPANPLVTPPHIKENYFLPAYA 280
 QY 63 -----GTTALAF---LVLMPLVH--GRNLFFRSLSWSWPPMLTLALAVILQNM 106
 DB 281 ILRSIPKMGSGVALARSIILFLPMLTHQAKQRTLAVERP-SQLMFWLVSTMLILT-- 337
 QY 107 AAHWVLETHDHPQLNRRVLYATFLF---PLNVLV 142
 DB 338 --WIGQPVH-HPYIQGIASITVPMFLALPLTALL 373

RESULT 15

Q37632 PRELIMINARY; PRT; 229 AA.
 ID Q37632;
 AC Q37632;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
 DE Cytochrome c oxidase polypeptide II (EC 1.9.3.1) (Fragment).
 GN COII.
 OS Rhagoletis alternata.
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pserygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Tephritidae; Tephritidae; Rhagoletis.
 OX NCBI_TaxID=47154;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97159559; PubMed=9007018;
 RA Smith J.J., Bush G.L.;
 RT "Phylogeny of the genus Rhagoletis (Diptera: Tephritidae) inferred
 RT from DNA sequences of mitochondrial cytochrome oxidase II.";
 RL Mol. Phylogenet. Evol. 7:33-43(1997).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINDING COPPER
 CC A CENTER TO THE BINETALIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 4 FERRICYTOCHROME C + O(2) = 4 FERRICYTOCHROME
 CC C + 2 H(2)O.
 CC -1- COPFACTOR: COPPER A AND HEME GROUP (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
 DR EMBL; U53260; AAB50353.1; -
 DR InterPro; IPR001505; Copper_CuA.
 DR InterPro; IPR002428; Cyt_c_ox_2.
 DR Pfam; PF00116; COX2.1.
 DR Pfam; PF02790; COX2_TM.1.
 DR PRINTS; PR01166; CYCOXIDASEII.
 DR PRODOM; PD000131; CopperCuA.1.
 DR PROSITE; PS00078; COX2_1.
 DR Copper; Inner membrane; Mitochondrion; Oxidoreductase;
 KW Respiratory chain; Transmembrane.
 FT NON TER 229
 SQ SEQUENCE 229 AA; 26559 MW; 2F99E26A33BEC148 CRC64;

Query Match 5.3%; Score 90.5; DB 8; Length 229;
 Best Local Similarity 22.0%; Pred. No. 3;
 Matches 63; Conservative 39; Mismatches 81; Indels 103; Gaps 14;

```

QY 37 SWNLFVPLPLPPLAGLLVQIIFP-----LGTALAFVLMPVLH---GRNLLFF 84
Db 3 TWN-----NIGLQDSASPLMEQLTFPHDHTLMTLWITTLVGLMFLFPNHYTNRLHG 58
QY 85 RSLSSWPFMTLALAVILQNMAMWVLETGHPQLTNRRVLYAATFLPLPLNVLVGA 144
Db 59 QTIEMIW-----TIIPAILLFLFA-----LPSRLLY-----LLEIN----- 91
QY 145 MXXXCPSI---AIRHPTPGVYTYRNFLKIB-----VSQSHPAMTAFCSLLLOAOSLLP- 195
Db 92 -----EPMTLKAIGHQWYTYEYSDFINVEFDSYMIQTNELTMDMFLLDVDRVILPM 146
QY 196 ---RTMAAPODSLRP-----GEEDEGM--QLQTKDSMAKARPAGXRGRAWGLAYT 243
Db 147 NSQIRILVTADVIHSMWTPISLGVKIDGTPGRINQTNFLM---NRPLFYGO----- 195
QY 244 LHNFTLOVFRKTALLGANGAOPCSSLPQSPSITPAMQAPGPPDH 289
Db 196 -----CSEICGANHSFMPVIESLPVNH 218

```

Search completed: March 13, 2003, 11:30:02
 Job time : 40.1126 secs

GenCore version 5.1.4.p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:26:57 ; Search time 10.5439 Seconds
(without alignments)
1278.447 Million cell updates/sec

Title: US-09-816-653A-4
Perfect score: 1695
Sequence: 1 CYSALVLSGLTFLVLMRS.....APPTWLMSQGFMRKLVG 325

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	94	5.5	476	TR16_HUMAN	Q15654 homo sapien
2	93	5.5	2016	CIN5_HUMAN	Q14524 homo sapien
3	91	5.4	346	RX_HUMAN	Q9Y2V3 homo sapien
4	90.5	5.3	625	NIFA_AZOB	P30667 azospirillum
5	89.5	5.3	395	YIGN_SALTY	O08022 salmoneilla
6	89	5.3	3164	TECU_HSV1	P10220 herpes simp
7	88.5	5.2	107	YGBE_ECOLI	P46141 escherichia
8	87	5.1	621	GRBA_MOUSE	Q60760 mus musculu
9	86	5.1	614	YDNK_LACLC	P42377 lactococcus
10	85.5	5.0	446	NY5R_PIG	O97969 sus scrofa
11	85.5	5.0	736	DVL2_HUMAN	O14641 homo sapien
12	84.5	5.0	1508	A10A_MOUSE	O54827 mus musculu
13	84	5.0	193	PMP2_MOUSE	P43925 mus musculu
14	84	5.0	321	NUJM_DINSE	O79546 dinodon sem
15	84	5.0	527	SPUB_SERMA	P21409 serratia ma
16	83	4.9	222	YJ73_SYNY3	P73933 synechocyst
17	83	4.9	1840	CIN4_BAT	P15390 rattus norv
18	82.5	4.9	228	COX2_ZOON	P29881 zoeteromopi
19	82.5	4.9	300	Y223_HAETN	P44579 haemophilus
20	82	4.8	347	NU2M_PANTR	O21798 pan troglod
21	81	4.8	335	YEB1_MYCAV	O07395 mycobacteri
22	81	4.8	637	LBR_CHICK	P23913 gallus gall
23	81	4.8	1499	A10C_HUMAN	O60312 homo sapien
24	80.5	4.7	1038	ATA3_MOUSE	O64518 mus musculu
25	80	4.7	329	YCIC_PSEPU	O47100 pseudomonas
26	80	4.7	379	CYB_ARTFI	O95728 artibeus fi
27	80	4.7	519	SCS9_YEAST	P20048 saccharomyc
28	80	4.7	654	SPH2_HUMAN	O9na48 homo sapien
29	80	4.7	859	M3KC_HUMAN	Q12852 homo sapien
30	80	4.7	988	M3KC_MOUSE	O60700 mus musculu
31	80	4.7	944	MMU3_MYCTU	O5357 mycobacteri
32	79.5	4.7	353	ILB8_MACMU	O28619 macaca mula
33	79.5	4.7	739	PURL_CAUCR	Q945f0 caulobacter

34	79	4.7	235	1	CD8A_HUMAN	P01732 homo sapien
35	79	4.7	1304	1	S3B1_HUMAN	O75533 homo sapien
36	79	4.7	1304	1	S3B1_MOUSE	O99nb9 mus musculu
37	79	4.7	1307	1	S3B1_XENLA	O57683 xenopus lae
38	78.5	4.6	511	1	MYIN_BUCAL	P57415 buchnera ap
39	78.5	4.6	721	1	ENP1_TORCA	P14400 torpedo cal
40	78	4.6	347	1	NU2M_BALPH	P24970 balanopter
41	78	4.6	351	1	KLF2_RAT	O9et58 rattus norv
42	78	4.6	1473	1	NAL1_HUMAN	O9c000 homo sapien
43	77.5	4.6	342	1	HUPK_AZOVI	P31878 azotobacter
44	77.5	4.6	342	1	RX_MOUSE	O35602 mus musculu
45	77.5	4.6	348	1	HOKV_AZOVI	P40597 azotobacter

ALIGNMENTS

RESULT 1
ID TR16_HUMAN STANDARD; PRT; 476 AA.
AC Q15654; Q15275; Q15170; Q9B7B2; Q9UN74;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Thyroid receptor interacting protein 6 (TRIP6) (OPA-interacting
protein 1) (Zyxin related protein 1) (ZRP-1).
GN TRIP6 OR OIP1.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98260683; PubMed=9598321;
RA Yi J., Backert M.C.;
RT "The human TRIP6 gene encodes a LIM domain protein and maps to
chromosome 7q22, a region associated with tumorigenesis.";
RL Genomics 49:314-316(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99329089; PubMed=10400701;
RA Murthy K.K., Clark K., Fortin Y., Shen S.-H., Banville D.;
RT "ZRP-1, a zyxin-related protein, interacts with the second PDZ domain
of the cytosolic protein tyrosine phosphatase hPTP1C.";
RL J. Biol. Chem. 274:20679-20687(1999).
RN [3]
RP SEQUENCE FROM N.A., AND VARIANT PHE-296.
RC TISSUE=Kidney, and Pancreas;
RA Straube R.;
RN Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RX [4]
RP SEQUENCE OF 310-476 FROM N.A.
RX MEDLINE=95295737; PubMed=776974;
RA Lee J.W., Choi H.-S., Gryulis J., Brent R., Moore D.D.;
RT "Two classes of proteins dependent on either the presence or absence
of thyroid hormone for interaction with the thyroid hormone
receptor.";
RL Mol. Endocrinol. 9:243-254(1995).
CC -|- SUBUNIT: Specifically interacts with the ligand binding domain of
the thyroid receptor (TR). Requires the presence of thyroid
hormone for its interaction. Interacts with PTPN13.
CC -|- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN KIDNEY, LIVER AND
LUNG. LOWER LEVELS IN HEART, PLACENTA AND PANCREAS.
CC -|- SIMILARITY: CONTAINS 3 LIM DOMAINS. THE LIM DOMAIN BINDS 2
ZINC IONS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
or send an email to license@ebi.ac.uk).

```

CC -----
DR EMBL: AF001902; CAA05080.1; -
DR EMBL: AF000974; AAB62222.1; -
DR EMBL: AF093834; AAD03037.1; -
DR EMBL: AF093834; AAD03037.1; JOINED.
DR EMBL: AF093835; AAD03037.1; JOINED.
DR EMBL: BC004249; AAH04249.1; -
DR EMBL: BC004999; AAH04999.1; -
DR EMBL: BC021540; AAH21540.1; -
DR EMBL: LA0374; AAC1740.1; -
DR Genew: HGNC:12311; TRIP6.
DR MIM: 602933; -
DR InterPro: IPR001781; LIM.
DR Pfam: PF00412; LIM; 3.
DR ProDom: PD000094; LIM; 3.
DR SMART: SM00132; LIM; 3.
DR PROSITE: PS00478; LIM_DOMAIN_1; 2.
DR PROSITE: PS50023; LIM_DOMAIN_2; 3.
DR LIM domain; Repeat; Metal-binding; Zinc; Polymorphism.
KM LIM domain; Repeat; Metal-binding; Zinc; Polymorphism.
FT DOMAIN 279 316 LIM 1.
FT DOMAIN 339 398 LIM 2.
FT DOMAIN 399 467 LIM 3.
FT VARIANT 296 296 L -> F.
FT CONFLICT 102 102 /FTID=VAR_013309.
FT CONFLICT 106 106 S -> T (IN REF. 2; AAB62222).
FT CONFLICT 135 135 S -> K (IN REF. 2; AAB62222).
FT CONFLICT 310 313 S -> C (IN REF. 1).
FT CONFLICT 461 476 CRAO -> MGP (IN REF. 4).
FT CONFLICT 461 476 SAMRLOEASATVTTTC -> RCGASRSSOPSGLTAESSMK
SQ SEQUENCE 476 AA, 50288 MW, 2847C747DD30A8FD CRC64;
Query Match 5.5%; Score 94; DB 1; Length 476;
Best Local Similarity 27.4%; Pred. No. 2.8;
Matches 43; Conservative 14; Mismatches 66; Indels 34; Gaps 7;
QY 175 GSHPMATATCSL-----LLOAOSLFRMAAPD--SLRGEEDGCM 214
DB 39 QHPFRN-FCPLPSECYCPGCPEDRGPAMVSGVLQHTGCLPADRGSLRGLDAEI 97
QY 215 QLLQTKDSMAKAGPAGXRGARWGLAYTLNPTLQVFRKALLGANGAOCSSLP--- 271
DB 98 DLLSLTLALNGRGHAKRRPDRQ-----YEPPEPPYRFGSLKPNPASPYPYGG 151
QY 272 GSPSITPAMQAPGPPDHXGVVCLHWPXXGSAP 308
DB 152 PTPASTYTAATPAGP---AFPVQVKVA-QPVRCGCP 184
RESULT 2
CINS_HUMAN STANDARD; PRT; 2016 AA.
ID CINS_HUMAN STANDARD; PRT; 2016 AA.
AC O14524;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sodium channel protein, cardiac muscle alpha-subunit (HML).
GN SCNSA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=92115699; PubMed=1309946;
RA Gallens M.E., George A.L. Jr., Chen L.Q., Chantline M., Horn R.,
RA Barchi R.L., Kallen R.G.;
RT "Primary structure and functional expression of the human cardiac
RT tetradotoxin-insensitive voltage-dependent sodium channel.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:554-558(1992).
RP [2]
RP VARIANTS LQ73.

```

```

RX MEDLINE=95196273; PubMed=7889574;
RA Wang Q., Shen J., Splawski I., Atkinson D., Li Z., Robinson J.L.,
RA Moss A.J., Towbin J.A., Keating M.T.;
RT "SCNSA mutations associated with an inherited cardiac arrhythmia, long
RT QT syndrome.";
RL Cell 80:805-811(1995).
RN [3]
RP VARIANTS LQ73.
RX MEDLINE=96081224; PubMed=8541846;
RA Wang Q., Shen J., Li Z., Timothy K., Vincent G.M., Priori S.G.,
RA Schwartz P.J., Keating M.T.;
RT "Cardiac sodium channel mutations in patients with long QT syndrome,
RT an inherited cardiac arrhythmia.";
RL Hum. Mol. Genet. 4:1603-1607(1995).
RN [4]
RP VARIANT LQ73 1505-LYS--GLN-1507 DEL.
RX MEDLINE=95379949; PubMed=7651517;
RA Bennett P.B., Yazawa K., Makita N., George A.L. Jr.;
RT "Molecular mechanism for an inherited cardiac arrhythmia.";
RL Nature 376:683-685(1995).
RN [5]
RP VARIANT LQ73 GLY-1790.
RX MEDLINE=98349542; PubMed=9686753;
RA An R.H., Wang X.L., Kerem B., Benhorin J., Medina A., Goldmit M.,
RA Kaas R.S.;
RT "Novel LQ7-3 mutation affects Na+ channel activity through
RT interactions between alpha- and beta-subunits.";
RL Circ. Res. 83:141-146(1998).
RN [6]
RP VARIANT LQ73 GLN-1623.
RX MEDLINE=98165676; PubMed=9506831;
RA Makita N., Shirai N., Nagashima M., Matsuoaka R., Yamada Y., Tohse N.,
RA Kitabatake A.;
RT "A de novo missense mutation of human cardiac Na(+) channel exhibiting
RT novel molecular mechanisms of long QT syndrome.";
RL EBBS Lett. 423:5-9(1998).
RN [7]
RP VARIANT LQ73 GLY-1839.
RX MEDLINE=98165676; PubMed=9506831;
RA Behorin J., Goldmit M., Maccluer J.W., Blangero J., Goffen R.,
RA Leibovitch A., Rahat A., Wang Q., Medina A., Towbin J., Kerem B.;
RT "Identification of a new SCNSA mutation, D1840G, associated with the
RT long QT syndrome.";
RL Hum. Mutat. 12:72-72(1998).
RN [8]
RP VARIANT LQ73 GLN-1623.
RX MEDLINE=98165676; PubMed=9506831;
RA Yamagishi H., Furutani M., Kamisago M., Morikawa Y., Kojima Y.,
RA Hino Y., Furutani Y., Kimura M., Imamura S.-I., Takao A., Momma K.,
RA Matsuoaka R.;
RT "A De Novo missense mutation (R1623Q) of the SCNSA gene in a Japanese
RT girl with sporadic long QT syndrome.";
RL Hum. Mutat. 12:481-481(1998).
RN [9]
RP VARIANTS LQ73 MET-1304 AND MET-1645, AND VARIANT ASN-1500.
RX MEDLINE=99439526; PubMed=10508990;
RA WatanabeRitschaigoon D., Vesely M.R., Duggal P., Levine J.C.,
RA Blume E.D., Wolff G.S., Edwards S.B., Beggs A.H.;
RT "Sodium channel abnormalities are infrequent in patients with long QT
RT syndrome: identification of two novel SCNSA mutations.";
RL Am. J. Med. Genet. 86:470-476(1999).
RN [10]
RP VARIANT LQ73 LYS-1784.
RX MEDLINE=99307063; PubMed=10377081;
RA Wei J., Wang D.W., Alings M., Fish F., Wachen M., Roden D.M.,
RA George A.L. Jr.;
RT "Congenital long-QT syndrome caused by a novel mutation in a conserved
RT acidic domain of the cardiac Na+ channel.";
RL Circulation 99:3165-3171(1999).
RN [11]
RP VARIANTS LQ73 N-1114; V-1501; L-1623; H-1644 AND N-1787.
RX MEDLINE=90432616; PubMed=10973849;
RA Splawski I., Shen J., Timothy K.W., Lehmann M.H., Priori S.,
RA Robinson J.L., Moss A.J., Schwartz P.J., Towbin J.A., Vincent G.M.,
RA Keating M.T.;

```

RT	"Spectrum of mutations in long-QT syndrome genes. KVLQTL, HERG, SCN5A, KCNE1, and KCNE2.";	
RL	Circulation 102:1176-1185(2000).	
CC	-I- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT SODIUM ION PERMEABILITY OF EXCITABLE MEMBRANES. ASSUMING OPENED OR CLOSED CONFORMATIONS IN RESPONSE TO THE VOLTAGE DIFFERENCE ACROSS THE MEMBRANE, THE PROTEIN FORMS A SODIUM-SELECTIVE CHANNEL THROUGH WHICH NA+ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL GRADIENT. IT IS A TETRODOTOXIN-RESISTANT NA+ CHANNEL ISOFORM. THIS CHANNEL IS RESPONSIBLE FOR THE INITIAL UPSTROKE OF THE ACTION POTENTIAL IN THE ELECTROCARDIOGRAM.	
CC	-I- SUBCELLULAR LOCATION: Integral membrane protein.	
CC	-I- TISSUE SPECIFICITY: EXPRESSED IN HUMAN ATRIAL, VENTRICULAR CARDIAC MUSCLE BUT NOT IN ADULT SKELETAL MUSCLE, BRAIN, MYOMETRIUM, LIVER, OR SPLEEN.	
CC	-I- DOMAIN: THE SEQUENCE CONTAINS 4 INTERNAL REPEATS, EACH WITH 5 HYDROPHOBIC SEGMENTS (S1,S2,S3,S5,S6) AND ONE POSITIVELY CHARGED SEGMENT (S4). SEGMENTS S4 ARE PROBABLY THE VOLTAGE-SENSORS AND ARE CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.	
CC	-I- DISEASE: DEFECTS IN SCN5A ARE THE CAUSE OF LONG QT SYNDROME TYPE 3 (LQT3), AN AUTOSOMAL DOMINANT CARDIAC DISEASE CHARACTERIZED BY RECURRENT SYNCOPE AND SUDDEN CARDIAC DEATH.	
CC	-I- MISCELLANEOUS: NA+ CHANNELS IN MAMMALIAN CARDIAC MEMBRANE HAVE FUNCTIONAL PROPERTIES QUITE DISTINCT FROM NA+ CHANNELS IN NERVE AND SKELETAL MUSCLE.	
CC	-I- SIMILARITY: TO OTHER SODIUM CHANNEL PROTEINS.	
CC	-I- SIMILARITY: CONTAINS 1 IQ DOMAIN.	
CC	-I- DATABASE: NAME=LOTSdb; NOTE=SCN5A mutations page; WWW="http://www.ssi.dk/en/forhning/qetdb/scn5a.htm".	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).	
CC	-----	
DR	EMBL: M77235; AAA58644.1; -	
DR	GeneW: HGNC:10593; SCN5A.	
DR	MIM: 600163; -	
DR	InterPro: IPR001682; Ca/Na_pore.	
DR	InterPro: IPR002111; Cat_channel_TpL.	
DR	InterPro: IPR000048; IQ_region.	
DR	InterPro: IPR000636; M+channel_nlg.	
DR	InterPro: IPR001696; Na_channel.	
DR	Pfam: PF00520; ion_trans; 4.	
DR	Pfam: PF00612; IQ; 1.	
DR	PRINTS: PR00170; NACHANNEL.	
DR	SMART: SM00015; IQ; 1.	
KM	Ionic channel; Transmembrane; Ion transport; Voltage-gated channel; Glycoprotein; Repeat; Multigene family; Phosphorylation;	
KM	Polymorphism; Disease mutation; Long QT syndrome	
FT	TRANSMEM 127 150	
FT	TRANSMEM 159 178	S1 OF REPEAT I (POTENTIAL).
FT	TRANSMEM 192 210	S2 OF REPEAT I (POTENTIAL).
FT	TRANSMEM 217 236	S3 OF REPEAT I (POTENTIAL).
FT	TRANSMEM 253 276	S4 OF REPEAT I (POTENTIAL).
FT	TRANSMEM 390 415	S5 OF REPEAT I (POTENTIAL).
FT	TRANSMEM 712 736	S6 OF REPEAT II (POTENTIAL).
FT	TRANSMEM 748 771	S1 OF REPEAT II (POTENTIAL).
FT	TRANSMEM 799 799	S2 OF REPEAT II (POTENTIAL).
FT	TRANSMEM 780 780	S3 OF REPEAT II (POTENTIAL).
FT	TRANSMEM 806 825	S4 OF REPEAT II (POTENTIAL).
FT	TRANSMEM 862 862	S5 OF REPEAT II (POTENTIAL).
FT	TRANSMEM 914 939	S6 OF REPEAT II (POTENTIAL).
FT	TRANSMEM 1201 1224	S1 OF REPEAT III (POTENTIAL).
FT	TRANSMEM 1238 1263	S2 OF REPEAT III (POTENTIAL).
FT	TRANSMEM 1270 1291	S3 OF REPEAT III (POTENTIAL).
FT	TRANSMEM 1296 1317	S4 OF REPEAT III (POTENTIAL).
FT	TRANSMEM 1337 1359	S5 OF REPEAT III (POTENTIAL).
FT	TRANSMEM 1444 1470	S6 OF REPEAT III (POTENTIAL).
FT	TRANSMEM 1524 1547	S1 OF REPEAT IV (POTENTIAL).

Query Match	Best Local Similarity	5.5%;	Score 93;	DB 1;	Length 2016;
Matches 67;	Conservative 31;	Mismatches 113;	Indels 70;	Gaps 13.	
FT TRANSMEM	1559	1582	S2 OF REPEAT IV (POTENTIAL).		
FT TRANSMEM	1589	1612	S3 OF REPEAT IV (POTENTIAL).		
FT TRANSMEM	1623	1644	S4 OF REPEAT IV (POTENTIAL).		
FT TRANSMEM	1660	1682	S5 OF REPEAT IV (POTENTIAL).		
FT TRANSMEM	1748	1772	S6 OF REPEAT IV (POTENTIAL).		
FT CARBOHYD	214	214	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	283	283	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	288	288	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	291	291	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	318	318	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	328	328	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	548	548	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	592	592	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	740	740	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	803	803	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	841	841	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	864	864	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	946	946	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	1365	1365	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	1374	1374	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	1380	1380	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	1388	1388	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	1736	1736	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	1774	1774	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	1955	1955	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT VARIANT	558	558	H -> R.		
FT VARIANT	1114	1114	/FtId=VAR_008955.		
FT VARIANT	1304	1304	D -> N (IN LOT3).		
FT VARIANT	1304	1304	/FtId=VAR_009935.		
FT VARIANT	1325	1325	T -> M (IN LOT3).		
FT VARIANT	1325	1325	/FtId=VAR_008956.		
FT VARIANT	1325	1325	N -> S (IN LOT3).		
Query Match	Best Local Similarity	5.5%;	Score 93;	DB 1;	Length 2016;
Matches 67;	Conservative 31;	Mismatches 113;	Indels 70;	Gaps 13.	
QY 35 QSSNLFSPPLPLPLAGLVGQIIIFPGITTLATLAVLMPVLHGNNLFFRSLESSMPPW 94	778 QQGMNIFD-----	LTATLAVILONMAAHV-----	FLTHDGHPOLTNR---	RVLY 129	
DB 826 NTLIKIGNSGALGNLTILVAILYIFPAVNGMOLEFKNYISLRSDGLPRMHMDPF 885	95 -----	LTATLAVILONMAAHV-----	FLTHDGHPOLTNR---	RVLY 129	
QY 130 AATFLFPLNLVGMWXXXCSPSIAIRHPTPGYTYRNFLEKIEVSQSHPMATAPCSLLQ 189	886 HAFLLIF--RIICGEMIEITMDMEVSGS--	LCLVFLTLMVIGNLVVLNLFALLLS 940			
QY 190 AOSLPRPTMAAQDSLRPEDEEGMQLLOTKRSMAGKAPGAXRGRARGLAVTYLL--HNP 248	941 SFS-----	ADNLTAPDEDERENNLIOL--	ALALIQGLRFPVKTKTTWDPCCGLRHRP 989		
QY 249 TLQVFRKTLALGANGAQP--CSSLPSSPSITPAMQACPPD 288	990 -----	OKPAAALAAQGLPSCIAIPIYSP-----	PPE 1015		
DB 990 -----	OKPAAALAAQGLPSCIAIPIYSP-----	PPE 1015			
RESULT 3					
RX HUMAN					
ID RX HUMAN	STANDARD;	PRT;	346 AA.		
AC Q9TYV3;					
DT 16-OCT-2001 (Rel. 40, Created)					
DT 15-OCT-2001 (Rel. 40, Last sequence update)					
DT 15-JUN-2002 (Rel. 41, Last annotation update)					
DE Retinal homeobox protein Rx (Retina and anterior neural fold homeobox protein).					
GN RAX OR RX.					
OS Homo sapiens (Human).					
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX NCBI_TaxId=9606;					
RN [1]					

RP SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RX MEDLINE=20092886; PubMed=10625658;
 RA Kimura A., Singh D., Mawousek E.F., Kikuchi M., Nakamura M.,
 RA Shinozaki T.,
 RT "Both PGE-1/RX and OXY/CRX interactions are necessary for
 RT photoreceptor-specific gene expression.";
 RT J. Biol. Chem. 275:1152-1160(2000).
 CC -1- FUNCTION: plays a critical role in eye formation by regulating the
 CC initial specification of retinal cells and/or their subsequent
 CC proliferation. Binds to the photoreceptor conserved element-1
 CC (PCE-1/Ret 1) in the photoreceptor cell-specific arrestin
 CC promoter.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: Expressed in the developing eye and weakly
 CC expressed in the adult retina.
 CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
 CC "BICOID" SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF15392; AAD2438.1; -.
 CC HSP; P06601; 1FUJ.
 CC Genew; HGNC:18662; RAX.
 DR MIM; 601881; -.
 DR InterPro; IPR003654; Homeo OAR.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 DR PROSITE; PS50803; OAR; 1.
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;
 KW Transcription regulation.
 FT DOMAIN 33 40 OCTAPEPTIDE MOTIF.
 FT DNA BIND 136 195 HOMEBOX.
 FT DOMAIN 323 336 OAR.
 FT DOMAIN 329 333 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 276 282 POLY-PRO.
 SQ SEQUENCE 346 AA; 36804 MW; 5F1D0512DCB90282 CRC64;
 Query Match 5.4%; Score 91; DB 1; Length 346;
 Best Local Similarity 23.8%; Pred. No. 3.4;
 Matches 48; Conservative 16; Mismatches 80; Indels 58; Gaps 6;

ID NIFA_AZOBR STANDARD; PRT; 625 AA.
 AC P30667;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Nif-specific regulatory protein.
 GN NIFA.
 OS Azospirillum brasilense.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
 OC Azospirillum.
 OC NCBI_TaxID=192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sp7;
 RX MEDLINE=92140038; PubMed=1779763;
 RA Liang Y.Y., Kaminski P.A., Elmerich C.;
 RT "Identification of a nifA-like regulatory gene of Azospirillum
 RT brasilense Sp7 expressed under conditions of nitrogen fixation and in
 RT the presence of air and ammonia.";
 RL Mol. Microbiol. 5:2735-2744(1991).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=9112522; PubMed=1362170;
 RA Liang Y.Y., de Zamaroczy M., Arsene F., Paquelin A., Elmerich C.;
 RT "Regulation of nitrogen fixation in Azospirillum brasilense Sp7:
 RT involvement of nifA, glnA and glnB gene products.";
 RL FEMS Microbiol. Lett. 79:113-119(1992).
 CC -1- FUNCTION: NIFA, A TRANSCRIPTIONAL ACTIVATOR, IS REQUIRED FOR
 CC ACTIVATION OF MOST NIF OPERONS, WHICH ARE DIRECTLY INVOLVED IN
 CC NITROGEN FIXATION. NIFA INTERACTS WITH SIGMA-54.
 CC -1- SIMILARITY: THE CENTRAL REGION CONTAINS A SIGMA-54 FACTOR
 CC INTERACTION ATP-BINDING DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X60714; CAA43126.1; -.
 CC PIR; S18420; S18420.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR001018; GAF.
 DR InterPro; IPR002197; HTH_Fis.
 DR InterPro; IPR002078; Sig54_Interact.
 DR Pfam; PF00158; Sigmas4_activat; 1.
 DR Pfam; PF01590; GAF; 1.
 DR SMART; SM00382; AAA; 1.
 DR SMART; SM00065; GAF; 1.
 DR TIGRfam; TIGR01199; HTH_fib; 1.
 DR PROSITE; PS00675; SIGMAS4_INTERACT_1; 1.
 DR PROSITE; PS00676; SIGMAS4_INTERACT_2; 1.
 DR PROSITE; PS00688; SIGMAS4_INTERACT_3; 1.
 DR PROSITE; PS50045; SIGMAS4_INTERACT_4; 1.
 KW Nitrogen fixation; Transcription regulation; Activator;
 KW ATP-binding; DNA-binding.
 FT DOMAIN 205 433
 FT DOMAIN 434 582 SIGMA-54 FACTOR INTERACTION (POTENTIAL).
 FT DOMAIN 583 625 INTER-DOMAIN LINKER.
 FT NP_BIND 233 240 C-TERMINAL DNA-BINDING DOMAIN.
 FT NP_BIND 296 305 ATP (POTENTIAL).
 FT METAL 447 447 ATP (POTENTIAL).
 FT METAL 447 447 BY SIMILARITY.
 FT METAL 452 452 BY SIMILARITY.
 FT DNA BIND 597 616 H-T-H MOTIF (BY SIMILARITY).
 SQ SEQUENCE 625 AA; 67855 MW; D060EA86C81FECC3 CRC64;
 Query Match 5.3%; Score 90.5; DB 1; Length 625;
 Best Local Similarity 27.7%; Pred. No. 7;
 Matches 48; Conservative 9; Mismatches 55; Indels 61; Gaps 9;

```

QY 184 CSLLLOASL--PRTMAPODSLRPGEDBGMQLLOTKDSMAKCA-----RPG-----230
D 447 CSLMNCSSVLFQYRTLCASVGLAP-----SMGCAINRVPDPRPGGPAA 492
QY 231 -----AXGRARWGLAYTLNHPPTLQVPRKTLALGANG-----AOP 266
D 493 ANAKTRPMAPVPEPAGAAAARGR-----ARRVPRPLGLRRRPA--GGSGPPDPACP 546
QY 267 CSSLPGSPSITPAMQAPGPDHGXVEVCLHWPXKSGAPPTLMESQCGFW 319
D 547 CPSPAPLPQAPPPSPAPAAPPP-----AAEVPLD-BPESGSLRDLMLMERTGW 595

RESULT 5
YJGN_SALTY STANDARD; PRT; 395 AA.
ID YJGN_SALTY STANDARD; PRT; 395 AA.
AC 008022;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein YJGN (ORF X).
GN YJGN OR STM4474.
OS Salmoneilla typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmoneilla.
OX NCBI_taxid=602;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
  Courtney L., Porcullik S., Ali J., Dante M., Du F., Hou S., Layman D.,
  Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
  Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
  Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmoneilla enterica serovar Typhimurium
  LT2."
RT Nature 413:852-856(2001).
RL (2)
RN SEQUENCE OF 1-315 FROM N.A.
RP STRAIN=LT2;
RX MEDLINE=94075212; PubMed=8253666;
RA Person B.C., Bjork G.R.;
RT "Isolation of the gene (wiae) encoding the hydroxylase involved in
  the synthesis of 2-methylthio-cis-ribozestatin in tRNA of Salmoneilla
  typhimurium and characterization of mutants."
RL J. Bacteriol. 175:7776-7785(1993).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE008910; AAL23293.1; -
DR EMBL: X73368; CAAS1785.1; -
DR PIR: S34364; S34364.
DR StGene; SG10443; YJGN.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 35 45
FT TRANSMEM 72 92
FT TRANSMEM 96 116
FT TRANSMEM 147 167
FT TRANSMEM 169 189
FT TRANSMEM 230 250
FT TRANSMEM 276 296
FT TRANSMEM 328 348
FT TRANSMEM 312 313
FT TRANSMEM 395 AA; 44107 MW; 04C59F293B731E2 CRC64;
SQ SEQUENCE

```

```

Query Match 5.3%; Score 89.5; DB 1; Length 395;
Best Local Similarity 26.2%; Pred. No. 5.1;
Matches 37; Conservative 20; Mismatches 45; Indels 39; Gaps 6;

QY 5 ALVSLCLLFLVLMNSLVTHRLGSGSGSDGQFSNLFVSFPLPLPPLAGLTVQOIIFFLGT 64
D 73 ALFISVLVLFVLIYVLSLIEHGHPLG-----FTLF-----GULL-ALLPFAV 116
QY 65 TALAFVLMFVHGRLLFFRSLSSESPFMTLTAVALIIONNAHWFLETHDGHPLTN 124
D 117 KGLQYQAMMTSLNGVHFQCSMRRAW--WYFALPVL--WVA----- 156
QY 125 RRVLYATFLPELVNVGAM 145
D 157 ---LYIVLYIISLVTVAVGGL 174

RESULT 6
TEGU_HSV11 STANDARD; PRT; 3164 AA.
ID TEGU_HSV11 STANDARD; PRT; 3164 AA.
AC P10250;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Large tegument protein (Virion protein UL36).
GN UL36.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_taxid=10299;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
  McNab D., Perry L.J., Scott J.E., Taylor P.;
RA "The complete DNA sequence of the long unique region in the genome of
  herpes simplex virus type 1."
RT J. Gen. Virol. 69:1531-1574(1988).
RL J. Gen. Virol. 69:1531-1574(1988).
CC -1- FUNCTION: TEGUMENT PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
  EBV-1 24, EBV BFLP1, HVS-1 64, VZV 22, AND HCMV UL48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X14112; CAJ3311.1; -
DR PIR: I30085; MMBE6.
DR InterPro; IPR005210; Herpes_UL36.
DR Pfam; PF03586; Herpes_UL36; 1.
KW Repeat.
FT DOMAIN 2911 2980
FT SEQUENCE 3164 AA; 335857 MW; C5D31FF4F9FE3F4 CRC64;

Query Match 5.3%; Score 89; DB 1; Length 3164;
Best Local Similarity 36.9%; Pred. No. 52;
Matches 24; Conservative 4; Mismatches 17; Indels 20; Gaps 3;

QY 247 NPTLQVFRKTLALGANGAOPCSSLPGSPSITPAMQAPGPDHGXVEVCLHWPXKSGA 306
D 2824 DPTAVV-----LGRNPAEPITSSPAGSPPPAPVAVAP-----FTSG-P 2863
QY 307 PPTNL 311
D 2864 PPTNL 2868

RESULT 7
YGBE_ECOLI

```

```

1D YGBE_ECOLI STANDARD; PRT; 107 AA.
AC P46141; Q46895;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein YGBE.
GN YGBE OR B2749.
OS Escherichia coli.
OC Bacteriophage.
OC Bacteriophage; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glanet S.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN (2)
RP SEQUENCE OF 1-81 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92268080; PubMed=1316900;
RA Lehn T.S., Vogt T.F., Suo Y.;
RT "The DNA sequence of the sulfate activation locus from Escherichia
RT coli K-12.";
RL J. Biol. Chem. 267:10405-10410(1992).
RN (3)
RP IDENTIFICATION;
RX MEDLINE=96032851; PubMed=7567469;
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Danchin A.;
RT "Detection of new genes in a bacterial genome using Markov models for
RT three gene classes.";
RL Nucleic Acids Res. 23:3554-3562(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U29579; AAC69259.1; -
DR EMBL; AE000358; AAC75791.1; -
DR EMBL; M74586; -; NOT ANNOTATED_CDS.
DR Ecogene; EGI2707; ygbE.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 67 81 VAAVVGIALHSLMDG -> SRSQGDCAAFIDGRK (IN
FT REF. 1).
SQ SEQUENCE 107 AA; 12017 MW; E3ED0F8E9EA9C21F CRC64;
Query Match 5.2%; Score 88.5; DB 1; Length 107;
Best Local Similarity 24.6%; Pred. No. 1.5;
Matches 28; Conservative 21; Mismatches 34; Indels 31; Gaps 5;
Oy 33 DQGFQSWNFSVPLPLPLAGLVQOIFFLGTTALFLVLPVL-HGRNLLFFRSLESSW 91
Db 18 DEETWS-----LPGAVV-----GRISWLFALAMPMLYGSNTLFF--FITW 58
Oy 92 PFWLTLAVALIIONNAAMWVLETHDGHPLQTNRRVLYAATFLLPPLVAVGM 145
Db 59 PFWLALMPVAVVGVIALH-----SLMDGKLRYSIIVFTLVTVGIMFGAL 101
RESULT 8
ID GRBA_MOUSE STANDARD; PRT; 621 AA.
AC Q60760; Q35352;
DT 15-JUL-1999 (Rel. 38, Created)

```

```

DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Growth factor receptor-bound protein 10 (GRB10 adaptor protein).
GN GRB10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss;
RX MEDLINE=95249278; PubMed=7731717;
RA Ooi J., Yajnik V., Immanuel D., Gordon M., Moskow J.J., Buchberg A.,
RA Margolis B.;
RT "The cloning of Grb10 reveals a new family of SH2 domain proteins.";
RL Oncogene 10:1621-1630(1995).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=97216049; PubMed=9062339;
RA Laviola L., Giorgino P., Chow J.C., Baquero J.A., Hansen H., Ooi J.,
RA Zhu J., Riedel H., Smith R.J.;
RT "The adapter protein Grb10 associates preferentially with the insulin
RT receptor as compared with the IGF-I receptor in mouse fibroblasts.";
RL J. Clin. Invest. 99:830-837(1997).
CC -1- FUNCTION: PLAYS A FUNCTIONAL ROLE IN INSULIN AND IGF-I SIGNALING.
CC MAY SERVE TO POSITIVELY LINK THE INSULIN AND IGF-I RECEPTORS TO AN
CC UNCHARACTERIZED MITOGENIC SIGNALING PATHWAY. INTERACTS WITH THE
CC CTOPPLASMIC DOMAIN OF THE AUTOPHOSPHORYLATED INSULIN RECEPTOR
CC WHICH IS THEN INHIBITED. THE INTERACTION IS MEDIATED BY THE SH2
CC DOMAIN. ALSO BINDS ACTIVATED PLATELET-DERIVED GROWTH FACTOR
CC RECEPTOR AND EPIDERMAL GROWTH FACTOR RECEPTOR.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE GRB7/10/14 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U8996; AAB51687.1; -
DR EMBL; AR022072; AAB72103.1; -
DR HSP; O60860; I012.
DR WGD; MG1:103232; Grb10.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000159; RA domain.
DR InterPro; IPR000980; SH2.
DR Pfam; PR00017; SH2.1.
DR Pfam; PR00169; PH.1.
DR Pfam; PR00788; RA.1.
DR PRINTS; PR00401; SH2DOMAIN.
DR ProDom; PD000093; SH2.1.
DR SMART; SM00233; PH.1.
DR SMART; SM00314; RA.1.
DR SMART; SM00252; SH2.1.
DR PROSITE; PSS0003; PH DOMAIN.1.
DR PROSITE; PSS0001; SH2.1.
KW SH2 domain; Alternative splicing.
FT DOMAIN 318 427 PH.
FT DOMAIN 520 601 SH2.
FT VARSPLIC 117 141 SH2.
FT CONFLICT 491 492 MISSING (IN ISOFORM 2).
FT CONFLICT 491 492 NG -> KR (IN REF. 2).
SQ SEQUENCE 621 AA; 70471 MW; 2A9A45D5842468A7 CRC64;
Query Match 5.1%; Score 87; DB 1; Length 621;
Best Local Similarity 21.7%; Pred. No. 13;
Matches 38; Conservative 28; Mismatches 59; Indels 50; Gaps 8;

```



```

QY 172 EVSOSHAPMTAFCSL-----LLO-----AQSLLPTMAPODSLPGEEDEG 213
DB 4 DINSVSLSNACMGQDTPALPLEGQGHASNGAASSSRGQPOASPRQMGSCVHI 63
QY 214 MOLLQOTDSNAKAKRPPAKRGRANWGLAYTLINPTLOVRKTKLLGANGAOPSSIRGS 273
DB 64 LRRLOEDDOOLRTA-----SLPAIPNPPPE-----LTGA-----APGS 96
QY 274 PPSTTPM-----QPAKPP-DHXGVVCLHMPKXGASAP-PTWLMSQGFWRKKL 323
DB 97 PPSVAFSSLPSPPPQPPAKHGRCEKMIPEGENTGNGKRTKIMRQPPFGQLSLTL 151

RESULT 9
YDNK_LACTC
ID YDNK_LACTC STANDARD; PRT; 614 AA.
AC P42377;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 70.0 kDa protein in dnaK 3' region (ORF4).
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MG1363;
RX MEDLINE=94172317; PubMed=8126443;
RA Eaton T.V., Shearman C.A., Gasson M.J.;
RT "Cloning and sequence analysis of the dnaK gene region of Lactococcus
RT lactis subsp. lactis";
RL J. Gen. Microbiol. 139:3253-3263(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X76642; CAA54090.1; -
DR Hypothetical protein.
SQ SEQUENCE 614 AA; 69869 MW; B1146136D09B89D9 CRC64;

Query Match
Best Local Similarity 19.9%; Score 86; DB 1; Length 614;
Matches 49; Conservative 35; Mismatches 74; Indels 88; Gaps 11;

QY 5 ALVLSCLLTFLVLMRSLVTHRLGSGSGSGDGFMSV-----PLPLPLAGLLVQOI 58
DB 332 SLIFSCFSVLVLMOUPL-----PWLKLFSTFTVQDPAKSTLFGLF----- 374
QY 59 IFFLGTALAPLVLMPLVHGRNLLFFRSLESSWPFMLTALAVALIIONMAHWVLETHDG 118
DB 375 -----SALSIVLILPIL-----LDKISGKTSYVLTGLIVPSIL-----G 410
QY 119 HPQLNLR-----RVLYAA-----TFLLPPLNVLVGNAXXXXCSPSIARHPTPGY 163
DB 411 FAEFRNRQKGSQPLFASQSLINKTPFNYMENPSIAIGEYL-----PQVIGSHQPYEK 466
QY 164 TVRNFLEKIE--VSOSHAPMTAFCSLLLAQASLLPTMAPODSLPGEEDEGMOLLQTD 221
DB 467 TIQQPKDKXNVGMNQAT-----VLSQRKLPESLA-----KSIQISD 506
QY 222 SMAKGA 227
DB 507 YSKKGS 512

```

```

AC 097969;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Neuropeptide Y receptor type 5 (NPY-5 receptor) (Y5
DE receptor)
GN NPY5R OR NPYR5.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Wraith A., Tornsten A., Chardon P., Harbicz I., Chowdhary B.P.,
RA Andersson L., Larhammar D.;
RT "Porcine NPY receptors NPY1R, NPY2R and NPY5R: cloning, mapping and
RT comparative analysis";
RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LMD; TISSUE=Kidney;
RA Ito Y., Minezawa M.;
RT "Sus scrofa gene for neuropeptide Y receptor type 5, complete cds.";
RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
CC - FUNCTION: RECEPTOR FOR NEUROPEPTIDE Y AND PEPTIDE YY. THE ACTIVITY
CC OF THIS RECEPTOR IS MEDIATED BY G PROTEINS THAT INHIBITS ADENYLATE
CC CYCLASE ACTIVITY. SEEMS TO BE ASSOCIATED WITH FOOD INTAKE. COULD
CC BE INVOLVED IN FEEDING DISORDERS (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: Integral membrane protein.
CC - SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC HIGHEST TO TACHYKININS RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF106083; AAD13778.1; -
DR EMBL; AB019185; BAA34055.1; -
DR InterPro; IPR000276; GPCR_Kinodpen.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODOPSN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; FALSE_NEG.
DR PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Phosphorylation; Lipoprotein; Palmitate.
FT DOMAIN 1 39
FT TRANSMEM 41 62
FT DOMAIN 63 74
FT TRANSMEM 75 95
FT DOMAIN 96 115
FT TRANSMEM 116 137
FT DOMAIN 138 157
FT TRANSMEM 158 178
FT DOMAIN 179 210
FT TRANSMEM 211 232
FT DOMAIN 233 311
FT TRANSMEM 312 391
FT DOMAIN 392 404
FT TRANSMEM 405 428
FT DOMAIN 429 446
FT CARBOHYD 10 17
FT CARBOHYD 17 17
FT DISULFID 114 198
FT LIPID 442 442
FT LIPID 442 442
SQ SEQUENCE 446 AA; 50474 MW; B7F616C2394C6C40 CRC64;

Query Match
Best Local Similarity 21.0%; Score 85.5; DB 1; Length 446;
Matches 69; Conservative 40; Mismatches 109; Indels 111; Gaps 16;

```

```

QY 58 IIFELGTTALAFVLMVHGRN-----LLFPRS-----LESSWP 92
DB 53 LIGFMGN-----LLILAAVMKRNOKTIVNPLIGNLAFSDLVLVLFCSPTLTSLVDOM 108
QY 93 F-----WTLALAVILQMAAHWFLETHDHPOLITNRVLYAAT 132
DB 109 FGKVMCHIMPLFQCVLVSTLILISIAIVRYHMKH-----PVSNLNLTHNGY 157
QY 133 FLFLPLVLVGMKXXXXXCSPIAIRHPPGYTY-----RNLKIEVSSSHA 179
DB 158 FLATVWTL-----GLAICSLPVPHSLVELQESFGSAMLSSRYLCVESWPSDSY 207
QY 180 MTAFCSLLOAQSLLP-----RTMA---APODSLRPEDEGMQL-LQTKDSM 223
DB 208 RIAFTISLLVLYILPLVCLTVSHTSVCRTISGGLSSQDSKL--EENEMINLTLP---- 261
QY 224 AKARFGAKRG-RARWGLATVTLHNPLOVFRKTLALGANGAOPC-SSLPGSPSITPM 281
DB 262 AKRSGQAKLSHHPKW--TYSFIRRRRRYSKKTACVRPAPALLESREGRPPKVGSM 319
QY 282 QPAGPPDH--XGXVEVCLHMEPYXGSAPP 308
DB 320 QSQPPPSNKMPCVPTCFEYKPEBNSDVP 348

```

```

RESULT 11
DVL2_HUMAN STANDARD; PRT; 736 AA.
ID 014641
AC 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Segment polarity protein dishevelled homolog DVL-2 (Dishevelled-2)
DE (DSH homolog 2).
GN DVL2
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=97336056; PubMed=9192851;
RA Semenov M.V., Snyder M.;
RT "Human dishevelled genes constitute a DHR-containing multigene
RT family.";
RL Genomics 42:302-310(1997).
CC -! FUNCTION: MAY PLAY A ROLE IN THE SIGNAL TRANSDUCTION PATHWAY
CC MEDIATED BY MULTIPLE NMT GENES.
CC -! SUBCELLULAR LOCATION: Cytoplasmic.
CC -! PTM: PHOSPHORYLATED.
CC -! SIMILARITY: BELONGS TO THE DSH FAMILY.
CC -! SIMILARITY: CONTAINS 1 DEP DOMAIN.
CC -! SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL; AF006012; AAB65243.1; -.
DR HSSP; Q12923; 3PDZ.
DR GENE; HGNC:3086; DVL2.
DR MIM; 602151; -.
DR InterPro; IPR000591; DEP.
DR InterPro; IPR001158; DIX.
DR InterPro; IPR003351; Dishevelled.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00610; DEP; 1.

```

```

DR Pfam; PF00778; DIX; 1.
DR Pfam; PF02377; Dishevelled; 1.
DR ProDom; PD003639; DIX; 1.
DR SMART; SM00021; DAX; 1.
DR SMART; SM00049; DEP; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50186; DEP; 1.
DR PROSITE; PS50106; PDZ; 1.
KM Developmental protein; Phosphorylation.
FT DOMAIN 7 12
FT DOMAIN 235 240 POLY-GLY.
FT DOMAIN 267 339 POLY-ARG.
FT DOMAIN 433 507 PDZ.
FT DOMAIN 686 694 DEP.
SQ SEQUENCE 736 AA; 78947 MW; 4BAD95B6C3FE531B CRC64;

```

```

Query Match 5.0%; Score 84.5; DB 1; Length 736;
Best Local Similarity 28.8%; Pred. No. 25;
Matches 36; Conservative 9; Mismatches 37; Indels 43; Gaps 6;

```

```

QY 204 SLRPEDEGMQLQTKDS--MAKARFGAKRGARWGL-----AYTLHNPLOVF 253
DB 630 SLRSGEASG-----TSDGPPPSRGSTGAPNLRAHGLHPYGPFGMALPVNPMVVM 684
QY 254 RKTALGANGAOPCSLPGSPSITPMOPAGPPDHGXVEVCLHMEPYXGSAPPTLWE 313
DB 685 -----MPPPPVPVPAVQPPGAPP-----VRDGSVPPE-LTA 716
QY 314 SQQGF 318
DB 717 SROGF 721

```

```

RESULT 12
A10A_MOUSE STANDARD; PRT; 1508 AA.
ID A10A_MOUSE
AC 054827
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Potential phospholipid-transporting ATPase VA (EC 3.6.3.1).
DE ATP10A OR ATPC5.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tetacarcinoma;
RX MEDLINE=20473714; PubMed=11015572;
RA Halleck M.S., Lawler J.F. Jr., Blackshaw S., Gao L., Nagarajan P.,
RA Hacker C., Pyle S., Newman J.T., Nakatani Y., Ando H., Weinstein D.,
RA Williamson P., Schlegel R.A.;
RT "Differential expression of putative transblayer amphipath
RT transporters.";
RL Physiol. Genomics 1:139-150(1999).
CC [2]
CC SEQUENCE OF 16-435 FROM N.A.
RX MEDLINE=98217376; PubMed=9548971;
RA Halleck M.S., Pradhan D., Blackman C.F., Berkes C., Williamson P.L.,
RA Schlegel R.A.;
RT "Multiple members of a third subfamily of P-type ATPases identified by
RT genomic sequences and ESTs.";
RL Genome Res. 8:354-361(1998).
CC -! CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC -! SUBCELLULAR LOCATION: Integral membrane protein.
CC -! TISSUE SPECIFICITY: FOUND IN TESTIS. ALSO DETECTED IN FETAL
CC TISSUES.
CC -! SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2
CC ATPASES). SUBFAMILY IV.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL, AF011337; AAC02902.1; -
 DR EMBL, AF156549; AF09447.1; -
 DR MGD; MGI:1330809; Atp10a.
 DR InterPro; IPR001757; Atpase_E1-E2.
 DR InterPro; IPR001454; Hlgase/hydrlase.
 DR Pfam; PF00702; Hydrolase.1.
 DR PRINTS; PR00119; CATAPPAE.
 DR PROSITE; PS00134; Atpase_E1_E2; 1.
 KM Hydrolyase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
 KM Multigene family.
 FT DOMAIN 1 79 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 80 101 POTENTIAL.
 FT TRANSSEM 102 107 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 108 129 POTENTIAL.
 FT TRANSSEM 130 313 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 314 335 POTENTIAL.
 FT TRANSSEM 336 366 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 367 388 POTENTIAL.
 FT TRANSSEM 389 1101 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 1102 1122 POTENTIAL.
 FT TRANSSEM 1123 1134 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 1135 1154 POTENTIAL.
 FT TRANSSEM 1155 1184 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 1185 1206 POTENTIAL.
 FT TRANSSEM 1207 1213 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 1214 1236 POTENTIAL.
 FT TRANSSEM 1237 1242 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 1243 1263 POTENTIAL.
 FT TRANSSEM 1264 1281 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 1282 1306 POTENTIAL.
 FT TRANSSEM 1307 1508 CYTOPLASMIC (POTENTIAL).
 FT MOD RES 431 431 PHOSPHORYLATION (BY SIMILARITY).
 FT METRL 1045 1045 MAGNESIUM (BY SIMILARITY).
 FT METRL 1049 1049 MAGNESIUM (BY SIMILARITY).
 FT DOMAIN 17 23 POLY-ARG.
 FT DOMAIN 471 474 POLY-ARG.
 FT CONFLICT 16 22 WRRPRR -> KLAARR (IN REF. 2).
 FT CONFLICT 435 435 T -> L (IN REF. 2).
 SQ SEQUENCE 1508 AA; 168699 MW; DC9A0D99A7EEB9E CRC64;

Query Match 5.0%; Score 84.5; DB 1; Length 1508;

Best Local Similarity 21.5%; Pred. No. 53;

Matches 53; Conservative 26; Mismatches 80; Indels 87; Gaps 11;

OY 59 IFFLTGTAFLVLMVPLHGRNLLPFRSLESSWPFVLTIALAVILQNMMAHWVLETHDG 118
 DB 1214 VFTWGPVTAIALFTPLH--LGIEITWTWVLAACG----- 1250
 OY 119 HPQLTNRVLYATFLLFPLNVLVGAMXXXCSPSIAIRHPTGYTYRNFLKEVQSHP 178
 DB 1251 -----STFLPFSVALIYNTSCATCP-----PSNPMYTMQTLL-----GDP 1286
 OY 179 AMTAFCSLLQAOGLPRT-MAAPQDSLRGEDEGMQLQ-----TKDSMAKGARP 229
 DB 1287 LFLYLTG-LIAPIALPLRLEFKALQSLPFTQLOLRQLAKPLNKSPDKSTFAQGGPP 1345
 OY 230 GAXRGARWGLAVTLNHTLQVFRKTAALGANGAOPCSLSPPSPS-----ITPAM 281
 DB 1346 G-----HSETELSERT--MG-----PFETLPKDCASQASQFTQOLTCS 1383
 OY 282 QPAGPP 287
 DB 1384 EASGEP 1389

RESULT 13
 PMP2_MOUSE

ID PMP2_MOUSE STANDARD; PRT; 193 AA.
 AC P42925;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 30-MAY-2000 (Rel. 35, Last annotation update)
 DE 22 kDa peroxisomal membrane protein.
 GN PMP2 OR PMP22.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss Webster; TISSUE=Liver;
 RX MEDLINE=96028630; PubMed=7551822;
 RA Bryant D.D., Wilson G.N.;
 RT "Differential evolution and expression of murine peroxisomal membrane
 RT protein genes";
 RL Biochem. Mol. Med. 55:22-30(1995).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Peroxisomal.
 CC -1- SIMILARITY: BELONGS TO THE MPV17 / PMP22 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL, L28835; AAA39957.1; -
 DR MGD; MGI:107487; Pmp2.
 KM Transmembrane; Peroxisome.
 FT INIT MET 0 0 BY SIMILARITY.
 FT TRANSSEM 1 29 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 30 50 POTENTIAL.
 FT TRANSSEM 51 73 PEROXISOMAL (POTENTIAL).
 FT TRANSSEM 74 94 POTENTIAL.
 FT TRANSSEM 95 112 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 113 133 POTENTIAL.
 FT TRANSSEM 134 171 PEROXISOMAL (POTENTIAL).
 FT TRANSSEM 172 192 POTENTIAL.
 SQ SEQUENCE 193 AA; 22134 MW; 493A467470381E0C CRC64;

Query Match 5.0%; Score 84; DB 1; Length 193;

Best Local Similarity 28.1%; Pred. No. 6.4;

Matches 34; Conservative 18; Mismatches 47; Indels 22; Gaps 4;

OY 9 SCLTFVLMRSIVTHRLSGSGSDGQFQSNLFSVPLPLPLGL-LVQOIFFLGTTA 66
 DB 62 SRLLEVSGLRLVLYGLFTVGPSHYLYLWETSV-PFVPMASVSKRLDRFFAPTL 120
 OY 67 LAFVLMVPLHGRNLL-----PFRSLESSWPFV-----LTIALAVILQNMMA 107
 DB 121 LFFPVNMLEGKQVSVFAVKMRSGFWPALQMWVMTPLQFININVPPLQFVLFANMA 180
 OY 108 A 108
 DB 181 A 181

RESULT 14
 NMIM_DINSE STANDARD; PRT; 321 AA.
 AC 079546;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE MADH-ubiquitinone oxidoreductase chain 1 (EC 1.6.5.3).
 GN MYND1 OR ND1.
 OS Dinodon semicarinatus (Akamata) (Dinodon semicarinatum).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Colubridae; Colubrinae; Dinodon.
OX NCBI_TaxID=56549;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA MEDLINE=98393589; PubMed=9725849;
RA Kumazawa Y., Ota H., Nishida M., Ozawa T.;
RT "The complete nucleotide sequence of a snake (Dinodon semicarinatus)
RL mitochondrial genome with two identical control regions.";
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB008539; BAA33022.1; -.
CC InterPro: IPR001694; Resp_NADH_dhl.
CC Pfam: PF00146; NADHdh_1.
CC DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.
CC DR PROSITE; PS00668; COMPLEX1_ND1_2; 1.
CC DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
CC SEQUENCE 321 AA; 36003 MW; 6907ECBAAC2FD9D CRC64;
SQ
Query Match
Best Local Similarity 5.0%; Score 84; DB 1; Length 321;
Matches 55; Conservative 29; Mismatches 76; Indels 110; Gaps 10;
QY 12 LTFVLVLRSLVTHR-LGSGGSGDGFQFVPLVPLPLAGLL-VQOIFELGTTALA 68
DB 108 LTFVAMSGMFTYALIMSGSSNSKY-----FLMGAMRAVAGIISYEVTGLI 155
QY 69 FLVLMVPLVGRNLLFFRS-----LESSWP--FWLTALA----- 100
DB 156 IISMAILSGYSIMLFTETOEHNMLLSSWPLAMWFTSLAETNRSFPLTEGSELSV 215
QY 101 -----VILQNMAMWVPLETHDGHQDLTRRVLYATFLLPLN 139
DB 216 GGNVRSAGPFLALFLAEYTNILFMNTLSTWML-----NPKMTNQ-----LFTIN 262
QY 140 VLVGAMXXXXCSPSIAIRHPTPGYTYRNPLKIEVSQHPMTAFCSLLQAQSLPRTMA 199
DB 263 LMTKTMLTT-----LFLMTRASYRFRPYDQMLMLWKQVLPPLTA 303
QY 200 APQDSLRLPGEDEGMQLLOTKDSMAKARP 229
DB 304 -----MYLNTSTSMALCGTP 319
RESULT 15
SFUB_SBRMA STANDARD; PRT; 527 AA.
AC IP21409;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Iron(III)-transport system permease protein sfub.
GN SFUB.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RX MEDLINE=90130286; PubMed=2404942;
RA Angerer A., Gaisner S., Braun V.;
RT "Nucleotide sequences of the sfuA, sfuB, and sfuC genes of Serratia

```

```

RT marcescens suggest a periplasmic-binding-protein-dependent iron
RT transport mechanism.";
RL J. Bacteriol. 172:572-578(1990).
CC -1- FUNCTION: INVOLVED IN A PERIPLASMIC BINDING-PROTEIN-DEPENDENT
CC IRON(III) TRANSPORT SYSTEM, WHICH IS INDEPENDENT OF TONB, EXB,
CC SIDEROPHORE, OR ANY OUTER MEMBRANE PROTEIN. HOWEVER, IT IS NOT
CC CLEAR HOW FE(III) IS SOLUBILIZED AND TAKEN UP ACROSS THE OUTER
CC MEMBRANE.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -1- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM PERMEASE FAMILY. CYSTW SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M33815; AAA26574.1; -.
CC PIR; B35108; QRSEUB.
CC InterPro: IPR000515; BPD_transp.
CC Pfam; PF00528; BPD_transp; 2.
CC DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; FALSE_NEG.
CC KM Transport; Iron transport; Transmembrane; Inner membrane; Repeat.
CC FT TRANSMEM 27 47
CC FT TRANSMEM 75 95
CC FT TRANSMEM 107 127
CC FT TRANSMEM 139 159
CC FT TRANSMEM 191 211
CC FT TRANSMEM 236 256
CC FT TRANSMEM 284 304
CC FT TRANSMEM 333 353
CC FT TRANSMEM 368 388
CC FT TRANSMEM 396 416
CC FT TRANSMEM 449 469
CC FT TRANSMEM 497 517
CC FT REPEAT 10 15
CC FT REPEAT 66 70
CC FT REPEAT 167 177
CC FT REPEAT 274 280
CC FT REPEAT 319 323
CC FT REPEAT 428 438
CC SQ SEQUENCE 527 AA; 56528 MW; 6B8154D100DB0504 CRC64;
Query Match
Best Local Similarity 5.0%; Score 84; DB 1; Length 527;
Matches 80; Conservative 30; Mismatches 133; Indels 102; Gaps 17;
QY 6 LVLSCLTFLVLMRSLVTHRLGSGGSGDGFQFVPLVPLPLAGLLVQOIFELGT 64
DB 243 LVLCCLGLLELLEISRGARVYAVGSGAR-----SCTRRLSPPLAL--ALLPLAL 294
QY 65 TALA---FLVLMVPLVGRNLLFFRSLEBSSWP--FWLTALAAILQNMAMWVPLETHDGH 119
DB 295 TALALGVPTTLARWMLGFGFVWRNAE-LWPLALMTLSLS----- 334
QY 120 POLTRNRVLYAATFLPLVNLVGAMXXXXCSPSIAIRHPT-----PGY 163
DB 335 -----AAGALLITICAIPMW-----LSVRYPARLYRVLEGNYTSSLPGLV 377
QY 164 TYRNPFLKIEVSQHPA-MTAFCSLLQAQSLPRTMAAPQDSLRLPGEDEGMQLLOTDS 222
DB 378 VALAVTITIHRSRPVYQVEITILLAYLMLPRALI-----NLRGIAQAPVELENVARS 433
QY 223 MARQ-----APGAKRGARWGLAYTLNPTLVQVFRKTLALGAGAPCCSL 270
DB 434 LGSPPAQALWSTTLRLAAGVNAAGAA--LVFLAIANE--LTAITLLAPRGTRTLAT- 485
QY 271 PGSPSITPAMQDPAGPPDXGKVENYCLHWEPRXKGSAPPTWMEQ 315

```

Thu Mar 13 14:11:15 2003

us-09-816-653a-4.rsp

Page 11

Db 486 -GFWALTSEIDYVAAP-YALIMVAL-----SPLTWLXSQ 520

Search completed: March 13, 2003, 11:30:26
Job time : 16.5439 secs

GenCore version 5.1.4.P5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:27:17 ; Search time 18.6069 Seconds

(without alignments)
1679.148 Million cell updates/sec

Title: US-09-816-653A-4

Perfect score: 1695

Sequence: 1 CYSALVLSCLTFLVLMRS.....APPTWLMESQGGFRKKLVG 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR:73:*\n2: p1r1:*\n3: p1r3:*\n4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96.5	5.7	593	2 A71361	probable thiamin A
2	93	5.5	2016	2 A38195	sodium channel pro
3	91.5	5.4	581	2 T44942	cytochrome-c oxida
4	90.5	5.3	625	2 S18420	regulatory protein
5	89.5	5.3	107	2 D85924	probable cytochrom
6	89.5	5.3	107	2 C91079	hypothetical prote
7	89.5	5.3	268	2 S34364	hypothetical prote
8	89	5.3	3164	1 WMBEH6	u36 protein - hum
9	88.5	5.2	107	2 A65086	hypothetical prote
10	88.5	5.2	395	2 AB1061	probable inner me
11	87	5.1	621	2 149159	growth factor rece
12	86.5	5.1	301	2 G83556	hypothetical prote
13	86	5.1	528	2 A10359	ion(III)-transpor
14	86	5.1	582	2 A87673	conserved hypoch
15	85	5.1	614	2 S40085	hypothetical prote
16	85.5	5.0	340	2 T18704	hypothetical prote
17	85.5	5.0	402	2 AB0854	probable membrane
18	84.5	5.0	166	2 T01263	hypothetical prote
19	84.5	5.0	230	2 T11290	cytochrome-c oxida
20	84.5	5.0	332	2 B81252	NADH2 dehydrogenas
21	84.5	5.0	377	2 F95948	probable acyltrans
22	84.5	5.0	636	2 T35182	probable ABC-type
23	84.5	5.0	1011	2 T07712	probable ABC-type
24	84	5.0	291	2 AG1846	sulfate transport
25	84	5.0	321	2 T11088	NADH2 dehydrogenas
26	84	5.0	477	2 D83617	probable amino aci
27	84	5.0	527	1 ORSEUB	stub protein - Ser
28	84	5.0	585	2 B5265	probable acid-CoA
29	83	4.9	222	1 S75137	hypothetical prote

30	83	4.9	297	2 AB2436	permease protein o
31	83	4.9	394	2 T31891	hypothetical prote
32	83	4.9	487	2 F72126	ct339 hypothetical
33	83	4.9	596	2 F75302	ATP-dependent DNA
34	83	4.9	1089	2 E82987	hypothetical prote
35	83	4.9	1840	1 CHRTM1	sodium channel pro
36	82.5	4.9	228	2 C38941	cytochrome-c oxida
37	82.5	4.9	393	2 B86644	transporter ybid l
38	82.5	4.9	888	2 JCS399	dual leucine zippe
39	82.5	4.9	3005	2 T08841	polypotein - dour
40	82	4.8	335	2 A87135	probable membrane
41	82	4.8	516	2 G98198	proline/betaine tr
42	82	4.8	516	2 AB3088	MPS permease [prol
43	82	4.8	627	2 A98344	hypothetical prote
44	82	4.8	627	2 AF2938	conserved hypochet
45	81.5	4.8	315	2 AG3535	dipeptide transpor

ALIGNMENTS

RESULT 1

A71361 Probable thiamin ABC transporter, permease protein - syphilis spirochete

C/Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C/Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999

C/Accession: A71361

R/Praser, C.M.; Norris, S.D.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin

rson, J.; Khailak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uettermack, T.; McDer

they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A/Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A/Reference number: A71250; MUID:98332770; PMID:9665876

A/Accession: A71361

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-593 <COL>

A/Cross-references: GB:AE001199; GB:AE000520; MID:G3322402; PIDN:AA65131.1; PID:G332240*

A/Experimental source: strain Nichols

C/Genetics:

A/Gene: TP0143

Query Match 5.7%; Score 96.5; DB 2; Length 593;
Best Local Similarity 23.0%; Pred. No. 1.9;
Matches 61; Conservative 29; Mismatches 86; Indels 89; Gaps 12;

QY	3	ISALVLSCLL-----TFLVMSLYTHRLGSGSGDGFSMVLFSVPLPLPLAGLVQOI	58
DB	120	LCSAVLCAVGFGAFFVKKSLRALFPA-----LCAIPLAIPPL-----	161
QY	59	IFPLGTALAFVIMPVLAGHNLFFRSLESSWPF-----WLTALA	100
DB	162	-----TALALFLSP-----GKNGICTRTLTGWSVTRPTLYSASGVVIAHAYNPLA	212
QY	101	VILGNMAHVFLETHDGHQPLTNR--RVLYAATFLLFPLNLVAGAMXXX-----	151
DB	213	LAIFARAWNTISADAEADARLLGARAVRFYTR-----LPALAGALRSSFLVFLYCF	267
QY	152	SIAL-----RHPTPGYTYRANFLKIEVSQSHP-----AMAFSGSLLL-----	188
DB	268	SLMWLLLGTTFTTLEVELEYS--IRTOAHFVASTALISQTFYALMLTMGVSTESQ	324
QY	189	QAOSLPRTMAPODSLRPGEDEG	213
DB	325	QAQSCVARTPLPPOSIGNIERIG	349

RESULT 2

A38195 sodium channel protein bH1, cardiac - human

N/Alternate names: tetrodotoxin-insensitive, voltage-dependent sodium channel, TTX-I NaCl

C/Species: Homo sapiens (man)

C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Aug-1999

Matches 28; Conservative 21; Mismatches 34; Indels 31; Gaps 5;

QY 33 DGOFSMNLFSVPLPLPPLAGLIVQOIIFFLGTTALAFVLMPVL-HGRNLLFFRSLESSW 91
 18 DEETWS-----LPGAVV-----GFISWLFALMPMLIYGSNTLFF--FIYTW 58

QY 92 PFMLTLAVALIIONMAAHWVLETHDGHPOLTNRVLYAATFLPLNVLVGAM 145
 59 PFFLALMPVAVVVGIALH-----SLMDGKLRSIVFTLVTGIMFGAL 101

Db 59 PFFLALMPVAVVVGIALH-----SLMDGKLRSIVFTLVTGIMFGAL 101

RESULT 6
 C91079
 probable cytochrome oxidase subunit [imported] - Escherichia coli (strain O157:H7, subsp
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: C91079
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
 A:Reference number: A9629; MUID:2156231; PMID:11258796
 A:Accession: C91079
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-107 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA837026.1; PID:g13363074; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RMD 0509952
 C:Genetics:
 A:Gene: EC63603

Query Match 5.3%; Score 89.5; DB 2; Length 107;
 Best Local Similarity 24.6%; Pred. No. 1.1;
 Matches 28; Conservative 21; Mismatches 34; Indels 31; Gaps 5;

QY 33 DGOFSMNLFSVPLPLPPLAGLIVQOIIFFLGTTALAFVLMPVL-HGRNLLFFRSLESSW 91
 18 DEETWS-----LPGAVV-----GFISWLFALMPMLIYGSNTLFF--FIYTW 58

QY 92 PFMLTLAVALIIONMAAHWVLETHDGHPOLTNRVLYAATFLPLNVLVGAM 145
 59 PFFLALMPVAVVVGIALH-----SLMDGKLRSIVFTLVTGIMFGAL 101

Db 59 PFFLALMPVAVVVGIALH-----SLMDGKLRSIVFTLVTGIMFGAL 101

RESULT 7
 S34364
 hypothetical protein X - Salmonella typhimurium (fragment)
 C:Species: Salmonella typhimurium
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
 C:Accession: S34364
 R:Perelson, B.C.; Bjork, G.R.
 submitted to the EMBL Data Library, June 1993
 A:Description: Isolation of a gene (miaE) involved in hydroxylation of ms2106A in trna c
 A:Reference number: S34359
 A:Accession: S34364
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-268 <PER>
 A:Cross-references: EMBL:X73368; NID:g312705; PIDN:CA51785.1; PID:g312711

Query Match 5.3%; Score 89.5; DB 2; Length 268;
 Best Local Similarity 26.2%; Pred. No. 3.1;
 Matches 37; Conservative 20; Mismatches 45; Indels 39; Gaps 6;

QY 5 ALVISCLEFLVLMRSIVTRILSGSGGQFSMNLFVPLPLPPLAGLIVQOIIFFLGTT 64
 27 ALFISVLLVFLIIVISLIEHGFPGI-----FTLE--GLLI--AIIFFMANV 70

QY 65 TALAFVLAMVVLGGRNLLFFRSLESSWPFMLTLAVALIIONMAAHWVLETHDGHPOLTN 124
 71 KGIQYQAMMTSLNGVHGFQCSKRRAM--MYMALPVL--MYA----- 110

Db 71 KGIQYQAMMTSLNGVHGFQCSKRRAM--MYMALPVL--MYA----- 110

QY 125 RRVLYAATFLPLNVLVGAM 145

Db 111 ---LYTVLIISLVIAVGCL 128

RESULT 8
 WMBE6
 UL36 protein - human herpesvirus 1 (strain 17)
 C:Species: human herpesvirus 1
 C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Jun-2000
 C:Accession: U13085
 R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Perry
 J. Gen. Virol. 69, 1531-1574, 1988
 A:Title: The complete DNA sequence of the long unique region in the genome of herpes sim
 A:Reference number: A30083; MUID:88274327; PMID:2839594
 A:Accession: U13085
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-3164 <MG>
 A:Cross-references: GB:X14112; NID:g1944536; PIDN:CA83231.1; PID:g59536; GB:D00317
 C:Genetics:
 A:Gene: UL36
 C:Superfamily: varicella-zoster virus gene 22 protein

Query Match 5.3%; Score 89; DB 1; Length 3164;
 Best Local Similarity 36.9%; Pred. No. 57;
 Matches 24; Conservative 4; Mismatches 17; Indels 20; Gaps 3;

QY 247 NPLQYFRKRTALGANGAQCSSLPGSPPSITPAMQAGPPDHGXVEVCLHMEPYXGSA 306
 2824 DPTAPV-----LGRNPABERTSSPAGSPPPAVQVAPF-----PTSG-P 2863

QY 307 PPTWL 311
 2864 PPTWL 2868

Db 2864 PPTWL 2868

RESULT 9
 A65056
 hypothetical protein in surF-cycC intergenic region - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: A65056
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coi
 .A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: A65056
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-107 <BLAT>
 A:Cross-references: GB:AE000358; GB:U00096; NID:g2367156; PIDN:AACT5791.1; PID:g1789106;
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: y90E

Query Match 5.2%; Score 88.5; DB 2; Length 107;
 Best Local Similarity 24.6%; Pred. No. 1.3;
 Matches 28; Conservative 21; Mismatches 34; Indels 31; Gaps 5;

QY 33 DGOFSMNLFSVPLPLPPLAGLIVQOIIFFLGTTALAFVLMPVL-HGRNLLFFRSLESSW 91
 18 DEETWS-----LPGAVV-----GFISWLFALMPMLIYGSNTLFF--FIYTW 58

QY 92 PFMLTLAVALIIONMAAHWVLETHDGHPOLTNRVLYAATFLPLNVLVGAM 145
 59 PFFLALMPVAVVVGIALH-----SLMDGKLRSIVFTLVTGIMFGAL 101

Db 59 PFFLALMPVAVVVGIALH-----SLMDGKLRSIVFTLVTGIMFGAL 101

RESULT 10
 AB1061
 probable inner membrane protein STY4013 [imported] - Salmonella enterica subsp. enteric
 C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called *Salmonella typhi*
 C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
 C/Accession: AB1061
 R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Conington, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A>Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov A:Reference number: AB0502; PMID:11677608
 A:Accession: AB1061
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-395 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD06935.1; PID:g15055583; GSPDB:GN00176 C/Genetics:
 A:Gene: SRY4813

Query Match
 Best Local Similarity 5.2%; Score 88.5; DB 2; Length 395;
 Matches 37; Conservative 20; Mismatches 45; Indels 39; Gaps 6;

Qy 5 ALVLSCLTFLVLMRSLVTHRLSGSGGCGQFSWNLFSVPLPLPLAGLLVQGIFFLGTT 64
 Db 73 ALFISVLVFTITVLSLSTFHEHPPGLG-----FTLF-----GLLI-ALIFPMV 116

Qy 65 TALAFVLMPVLHGRMLFFRSLSSMPFWLTALAVILONMAAHVPLETHDGHQQLTN 124
 Db 117 KGLQYQAMWMTSLNGVHFGFCSSMRAM--WYMLPVL--MYA----- 156

Qy 125 RRVLYAATPLPLPLVNLVGM 145
 Db 157 ---LVIVLYIISLVITAVGL 174

RESULT 11
 149199
 growth factor receptor binding protein Grb10 - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
 C/Accession: 149199
 R/Ooi, J.; Yajnik, V.; Immanuel, D.; Gordon, M.; Moskow, J.J.; Buchberg, A.M.; Margolis, Oncogene 10, 1621-1630, 1995
 A>Title: The cloning of Grb10 reveals a new family of SH2 domain proteins.
 A:Reference number: 149199; MUID:95249278; PMID:7731717
 A:Accession: 149199
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-621 <RES>
 A:Cross-references: EMBL:U18996; NID:g841209; PIDN:AA853687.1; PID:g841210 C/Genetics:
 A:Gene: Grb10
 C:Superfamily: pleckstrin repeat homology; SH2 homology
 C/Keywords: growth factor receptor
 F/520-614/Domain: SH2 homology <SH2B>

Query Match
 Best Local Similarity 5.1%; Score 87; DB 2; Length 621;
 Matches 38; Conservative 28; Mismatches 59; Indels 50; Gaps 8;

Qy 172 EVSOSHPRMTAFCSL-----LHQ-----AQSLPRMTAARQDSLRCGEDEG 213
 Db 4 DINSVEELNACNNOSTDTAPLEDDQAHNSGCASSSGQPOASFRQKQSSQPVHI 63

Qy 214 MQLIQTKSMKAPGAPGAKRGARWGLVYTLHNPTLVFPRKTLTGANGAQPCSSLPQS 273
 Db 64 LRRLOEEDQQLRTA-----SLPAIPNPFPE-----LTGA-----APGS 96

Qy 274 PPSITPAM---QPAQRP-DHXGXVEVCLHMEPPXGSA-PWLMESQCGFERKKL 323
 Db 97 PPSVAPSLPPPPQAPAKHGCRCWKIPGENTRGNKRIKRWQFPFGOLSKL 151

RESULT 12
 683556
 hypothetical protein PA0702 [imported] - *Pseudomonas aeruginosa* (strain PA01)
 C/Species: *Pseudomonas aeruginosa*
 C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C/Accession: G63556
 R/Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laribig, K.; Lim, .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho A:Reference number: AB2950; MUID:20437337; PMID:10584043
 A:Accession: G63556
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-301 <STO>
 A:Cross-references: GB:AE004506; GB:AE004091; NID:g9946584; PIDN:AA04091.1; GSPDB:GN001 A:Experimental source: strain PA01
 C/Genetics:
 A:Gene: PA0702

Query Match
 Best Local Similarity 5.1%; Score 86.5; DB 2; Length 301;
 Matches 61; Conservative 27; Mismatches 105; Indels 85; Gaps 10;

Qy 58 IIFPGLTALAFV-----LMPVHGR-----NLL----- 82
 Db 36 VALFLGALGSLFALERLPYQPRWRPHADRLDVLHASVNESLALGILALPLAGVLG 95

Qy 83 FFRSLSSMPFWLTALAVILONMAAHVPLETHDGHQQLTNRVLYAATFLPLNLV 142
 Db 96 FWPVWQAWPLMLQDLALVADLGTIVHYASH-----RSALLMRH-- 138

Qy 143 GAMXXCSPSIAIRHPTPGYTYRNFLKIEVSQHPMTAFCSLLQAQSLPRTMAAPQ 202
 Db 139 -----AVHNSVQRLYGFGLMKHPLHGLEALGTLPLLLG---VPQVALL 184

Qy 203 DSLRGEDEGMOLQTKDSMAKAPGAPGAKRGARWGLVYTLH-----NPTLVQV 253
 Db 185 AFA-----IGIQDLQHSNV--DMRIGLRHVFAPARPLRHIRYGRADVNFALFIS 236

Qy 254 RKTALGANGAOPCSSLPSPSITPAMOPACGPDHXG 291
 Db 237 VMDRLGCTALHRPDYRLDSTDGMI--GDQPDYRDYAG 272

RESULT 13
 A10359
 Iron(III)-transport system permease sfub [imported] - *Yersinia pestis* (strain CO92)
 C/Species: *Yersinia pestis*
 C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
 C/Accession: A10359
 R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.; demo-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davies, P.; Dougan, G.; i ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, Nature 413, 523-527, 2001
 A>Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: A10359
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-528 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC92204.1; PID:g15980916; GSPDB:GN00175 C/Genetics:
 A:Gene: sfub
 C:Superfamily: sfub protease

Query Match
 Best Local Similarity 5.1%; Score 86; DB 2; Length 528;
 Matches 78; Conservative 31; Mismatches 129; Indels 112; Gaps 15;

Qy 6 LVLSCLTFLVLMRSLVTHRLSGSGGCGQFSWNLFSVPLPLPLAGLLVQGIFFLGTT 65

Db 244 LALCCLGLLVEATSRGTRYARVSGTFR-----RQTVYSMTS 283
Qy 66 ALAFVLMFVLHGRNLLFFRSLESSWPF-----MLT-----LAL 99
Db 284 LFLCCLLPPL-----ITLSTSGVPTITLMRLMSIGIDIMWLPPELLAKOTLGAL 336
Qy 100 --AVILQMAAHWFLETHDHPOLTNRRVLYAATFLPLAVLVGAMXXXCSSIAIRH 157
Db 337 SGAVIITLCALPMWMLSR--YFGRHLR---AMEGCYVVSLSPLGIVALALVITIRI 390
Qy 158 PTGYYTYRNLKTEVSGSHPMATFCSLLQAOSLLPRTMAAPQDSLPGEDDEGMQL 217
Db 391 ARPLY-----QTEFTVLLAYLMLFTPRALI---SLRAGIAQAPVELE 429
Qy 218 QTKDSMAKG-----ARPGAKRGARWGLAYTLNHPLOVPRKTAALGANGAQ 265
Db 430 NVASISLGTPTQAMLSTTLRLAAGAAAGAA---LVFLAIGNE---LTATLLAPNCTR 482
Qy 266 PCSSLPSPSPITPAMOPAGPPDHGXVEVCLHMEPPXGSAAPTMYLMSQ 315
Db 483 TLAT--GFWALTSEIDYVAAP--YAFIMVAL-----SLPLTWLAYSQ 521

RESULT 14
A87673
converted hypothetical protein CC3419 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: A87673
R:Neuman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: A87673
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-582 <STO>
A:Cross-references: GB:AE005673; NID:g13425133; PIDN:AK25381.1; GSPDB:GN00148
A:Gene: CC3419

Query March 5.1%; Score 86; DB 2; Length 582;
Best Local Similarity 24.3%; Pred. No. 15;
Matches 74; Conservative 30; Mismatches 113; Indels 88; Gaps 14;

Qy 62 LGTTALAFVLMFVLHGRNLLFFRSLESSWPFMTLALAVILQMAAHWFLETHDG 118
Db 220 VGLTTLGWLPLMLVLIQGGFLATLPTGTGLAPLMGTAKATALLLSAAALILINAAODG 279
Qy 119 H--PQLTNRVLYAATFLPLAVLVG-----AMXXXCSP 151
Db 280 EBPPLILCLARIRGLILIPVILAGYALWLRIDGYLTPERVVAGYVLVAIGFTAGY 339
Qy 152 SIARHPTPGYTYRNLKTEVSGSHPMATFCSLL-----QAOSLL 194
Db 340 ALAAVKPGR-----WMK-PLERTNPITMAACVILLIALFTPIANPARLSVASQVKRL 391
Qy 195 PRTMAAPQ-----DSLPGEDDEGMQLQTKDS--MAKARPGAKRGARWGLAYTL 244
Db 392 SGEVAADKFDFOFLRFDAGRYGR--EALDRLKTHPNAETAKRADAA-----ASTE 440
Qy 245 LHNPTLOV---PRKTAALGANGAOPCS-----SLPGSPSPITPAMQ--PA--GPPDHGX 292
Db 441 KQYRGEIRPDFAKAMAVYPAGALPOSFVADWSQSPGSGSNTAIAIMQCPALVADVADGK 500
Qy 293 VEVCL 297
Db 501 DEVL 505

RESULT 15

S40085
hypothetical protein 4 - Lactococcus lactis
C:Species: Lactococcus lactis
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Oct-1999
C:Accession: S40085
R:Eaton, T.J.; Shearman, C.; Gasson, M.
submitted to the EMBL Data Library, December 1993
A:Description: Cloning and sequence analysis of the DNA gene region of Lactococcus lact
A:Reference number: S40085
A:Accession: S40085
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-614 <EAT>
A:Cross-references: EMBL:X76642; NID:g435489; PIDN:CAA54090.1; PID:g435493

Query March 5.1%; Score 86; DB 2; Length 614;
Best Local Similarity 19.9%; Pred. No. 16;
Matches 49; Conservative 35; Mismatches 74; Indels 88; Gaps 11;

Qy 5 ALVLSCLTFLVLMRSLVTHRLGSGSGDGFQSWNLFSV-----PLPLPPLAGLVQOI 58
Db 332 SLIFSCFSLVVLMLQPLI-----FPWKLFSIFIVQDPAFRTFLGLF--- 374
Qy 59 IFFLTALAFVLMFVLHGRNLLFFRSLESSWPFMTLALAVILQMAAHWFLETHDG 118
Db 375 -----SALSLVILPLI-----LDKISGKTSYVLTGLVIFSLI-----G 410
Qy 119 HPQLTNR-----RVLYAA-----TFLLPPLAVLVGAMXXXCSSIAIRHPTPGY 163
Db 411 FAERPRNRKQKSQSLPFAAOSLLNKTPFNWENDSIAIGYL---PQVIGSHNQPYEK 466
Qy 164 TYRNLKIE--VSGSHPMATFCSLLQAOSLLPRTMAAPQDSLPGEDDEGMQLQTKD 221
Db 467 TIQGFYDKVNYGWRNQAMT-----YLSRGKLPREGLA-----KSIQISD 506
Qy 222 SMAKGA 227
Db 507 YSKXGS 512

Search completed: March 13, 2003, 11:31:03
Job time : 23.6069 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:30:32 ; Search time 13.645 Seconds

(Without alignments)
1097.823 Million cell updates/sec

Title: US-09-816-653A-4

Perfect score: 1695
Sequence: 1 CYSALVLSGLTFLVLMRS.....APPTWLMSQGFMRKLVG 325Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications MA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEM_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/US06_NEM_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEM_PUB pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEM_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1679	99.1	325	US-09-816-653A-4	Sequence 4, Appli
2	1110.5	65.5	667	US-09-759-056-2	Sequence 2, Appli
3	1110.5	65.5	667	US-09-901-812-2	Sequence 2, Appli
4	1110.5	65.5	667	US-10-227-884-80	Sequence 80, Appli
5	1110.5	65.5	667	US-10-230-163-80	Sequence 80, Appli
6	1110.5	65.5	667	US-10-218-631-80	Sequence 80, Appli
7	1110.5	65.5	667	US-10-230-338-80	Sequence 80, Appli
8	1106.5	65.3	658	US-09-759-056-5	Sequence 5, Appli
9	1106.5	65.3	658	US-09-901-812-5	Sequence 5, Appli
10	832.5	49.1	670	US-09-816-653A-7	Sequence 7, Appli
11	485	28.6	178	US-09-825-299-796	Sequence 796, App
12	485	28.6	178	US-09-925-299-796	Sequence 796, App
13	97.5	5.8	403	US-09-826-508-30	Sequence 30, Appli
14	97.5	5.8	403	US-09-895-686-5	Sequence 5, Appli
15	97.5	5.8	427	US-09-826-508-32	Sequence 32, Appli
16	95.5	5.6	526	US-09-738-626-5712	Sequence 5712, Ap
17	93	5.5	2016	US-09-840-125-4	Sequence 4, Appli
18	92	5.4	664	US-09-767-870-9	Sequence 9, Appli
19	92	5.4	2100	US-09-995-542-6	Sequence 6, Appli

20	92	5.4	2144	10	US-09-858-194-2	Sequence 2, Appli
21	92	5.4	2146	10	US-09-995-542-5	Sequence 5, Appli
22	91.5	5.4	572	10	US-09-919-781-2	Sequence 2, Appli
23	89	5.3	249	9	US-09-738-626-6183	Sequence 6183, Ap
24	87	5.1	621	9	US-10-242-332-4	Sequence 4, Appli
25	87	5.1	2121	10	US-09-995-542-3	Sequence 3, Appli
26	87	5.1	2167	10	US-09-995-542-2	Sequence 2, Appli
27	84.5	5.0	299	9	US-09-510-332-7	Sequence 7, Appli
28	84.5	5.0	299	10	US-09-393-634-41	Sequence 41, Appli
29	84.5	5.0	1508	9	US-10-024-623-35	Sequence 35, Appli
30	84.5	5.0	1508	12	US-10-002-769-15	Sequence 15, Appli
31	84	5.0	686	9	US-10-174-590-524	Sequence 524, App
32	84	5.0	686	9	US-10-176-758-524	Sequence 524, App
33	84	5.0	686	9	US-10-175-737-524	Sequence 524, App
34	84	5.0	686	9	US-10-173-706-524	Sequence 524, App
35	84	5.0	686	9	US-10-175-738-524	Sequence 524, App
36	84	5.0	686	9	US-10-175-752-524	Sequence 524, App
37	84	5.0	686	9	US-10-176-482-524	Sequence 524, App
38	84	5.0	686	9	US-10-176-757-524	Sequence 524, App
39	84	5.0	686	9	US-10-176-913-524	Sequence 524, App
40	84	5.0	686	9	US-10-180-552-524	Sequence 524, App
41	84	5.0	686	9	US-10-180-557-524	Sequence 524, App
42	84	5.0	686	9	US-10-173-700-524	Sequence 524, App
43	84	5.0	686	9	US-10-174-572-524	Sequence 524, App
44	84	5.0	686	9	US-10-174-579-524	Sequence 524, App
45	84	5.0	686	9	US-10-174-582-524	Sequence 524, App

ALIGNMENTS

RESULT 1
US-09-816-653A-4
Sequence 4, Application US/0981653A
Publication No. US20030021788A1
GENERAL INFORMATION:
APPLICANT: Kaelelli, Luca
TITLE OF INVENTION: NOVEL HUMAN STRA6-LIKE PROTEIN AND NUCLEIC ACIDS ENCODING THE SAME
FILE REFERENCE: 10716/57
CURRENT APPLICATION NUMBER: US/09/816,653A
CURRENT FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: 60/191,532
PRIOR FILING DATE: 2000-03-23
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 325
TYPE: PRT
ORGANISM: Homo sapiens hSTRA6 polypeptide fragment, carboxy terminus
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (146)..(148)
OTHER INFORMATION: Xaa represents any amino acid, other, or unknown
NAME/KEY: MISC FEATURE
LOCATION: (232)..(232)
OTHER INFORMATION: Xaa represents any amino acid, other, or unknown
NAME/KEY: MISC FEATURE
LOCATION: (290)..(290)
OTHER INFORMATION: Xaa represents any amino acid, other, or unknown
NAME/KEY: MISC FEATURE
LOCATION: (292)..(292)
OTHER INFORMATION: Xaa represents any amino acid, other, or unknown
NAME/KEY: MISC FEATURE
LOCATION: (302)..(302)
OTHER INFORMATION: Xaa represents any amino acid, other, or unknown
NAME/KEY: MISC FEATURE
LOCATION: (303)..(303)
OTHER INFORMATION: Xaa represents any amino acid, other, or unknown
US-09-816-653A-4
Query Match 99.1%; Score 1679; DB 9; Length 325;
Best Local Similarity 100.0%; Pred. No. 7.6e-143;

APPLICANT: Desnoyers, Luc
APPLICANT: Gerltsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Watanabe, Colin L.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3530P1C79
CURRENT APPLICATION NUMBER: US/10/227,884
CURRENT FILING DATE: 2002-08-26
PRIOR APPLICATION NUMBER: 10/119,480
PRIOR FILING DATE: 2002-04-09
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063549
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/069873
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089905
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090691
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/095302
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/095318
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/095916
PRIOR FILING DATE: 1998-08-10
PRIOR APPLICATION NUMBER: 60/096146
PRIOR FILING DATE: 1998-08-11

PRIOR APPLICATION NUMBER: 60/096791
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: 60/097986
PRIOR FILING DATE: 1998-08-26
PRIOR APPLICATION NUMBER: 60/098544
PRIOR FILING DATE: 1998-08-31
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099803
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099811
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100038
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100848
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101741
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101786
PRIOR FILING DATE: 1998-09-25
PRIOR APPLICATION NUMBER: 60/101916
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101922
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/106178
PRIOR FILING DATE: 1998-10-28
PRIOR APPLICATION NUMBER: 60/106248
PRIOR FILING DATE: 1998-10-29
PRIOR APPLICATION NUMBER: 60/106464
PRIOR FILING DATE: 1998-10-30
PRIOR APPLICATION NUMBER: 60/106905
PRIOR FILING DATE: 1998-11-03
PRIOR APPLICATION NUMBER: 60/108787
PRIOR FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: 60/108801
PRIOR FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: 60/108849
PRIOR FILING DATE: 1998-11-18
PRIOR APPLICATION NUMBER: 60/112422
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: 60/113296
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/113605
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113621
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/115558
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115565
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115733
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/119549
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/123618

```

PRIOR FILING DATE: 1999-03-10
PRIOR APPLICATION NUMBER: 60/125259
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: 60/125775
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/126773
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: 60/127887
PRIOR FILING DATE: 1999-04-05
PRIOR APPLICATION NUMBER: 60/130232
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/131022
PRIOR FILING DATE: 1999-04-26
PRIOR APPLICATION NUMBER: 60/131270
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131291
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131445
PRIOR FILING DATE: 1999-04-28
PRIOR APPLICATION NUMBER: 60/134287
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/140650
PRIOR FILING DATE: 1999-06-22
PRIOR APPLICATION NUMBER: 60/140723
PRIOR FILING DATE: 1999-06-22
PRIOR APPLICATION NUMBER: 60/141037
PRIOR FILING DATE: 1999-06-23
PRIOR APPLICATION NUMBER: 60/144758
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/145698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 60/146222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: 60/146963
PRIOR FILING DATE: 1999-08-03
PRIOR APPLICATION NUMBER: 60/149320
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/149638
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/151733
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/164418
PRIOR FILING DATE: 1999-11-09
PRIOR APPLICATION NUMBER: 60/166361
PRIOR FILING DATE: 1999-11-16
PRIOR APPLICATION NUMBER: 60/169445
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: 60/169495
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: 60/169835
Query Match 65.5%; Score 110.5; DB 9; Length 667;
Best Local Similarity 76.6%; Pred. No. 1.6e-91;
Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;
QY 1 CYSALVLSCLTFLVLMRSIVTR-----LGSQSGS-----GQSWNLFS 42
DB 365 CYSALVLSCLTFLVLMRSIVTRHNTLRALHRAALDLSPLHRSPPSRQALFCQMSFS 424
QY 43 VPLPLPPLAGLIVVOOIFFLGTALAFVLMPLVHGRVLFRRSLSSSWPMLTALAVI 102
DB 425 AVQTAFLCLGLVVOOIFFLGTALAFVLMPLVHGRVLFRRSLSSSWPMLTALAVI 484
QY 103 LONMAAHVPLETHDGHQOLTNRRVLYATFLPLVNLVGMXXXC----- 149
DB 485 LONMAAHVPLETHDGHQOLTNRRVLYATFLPLVNLVGMXXXC----- 149
QY 150 -SPSLAIRP-----TGYTYRRNPFLKIEVSQSPHMTAFSCSLLOQSLPRTMAAPD 203
DB 545 GQMLSLPPLPAATLDDEGYTYRRNPFLKIEVSQSPHMTAFSCSLLOQSLPRTMAAPD 604
QY 204 SLRPEDEGQMLLOTXDSMAKGARPGASRGARMGLAYTLHNPPTLOVPRKTLALGANG 263

```

```

DB 605 SLRPEDEGQMLLOTXDSMAKGARPGASRGARMGLAYTLHNPPTLOVPRKTLALGANG 664
QY 264 AQP 266
DB 665 AQP 667
RESULT 5
US-10-230-163-80
Sequence 80, Application US/10230163
Publication No. US2003003635A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Desnoyers, Luc
APPLICANT: Gerritsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Watanabe, Colin L.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3510PIC96
CURRENT APPLICATION NUMBER: US/10/230.163
CURRENT FILING DATE: 2002-08-28
PRIOR APPLICATION NUMBER: 10/119.480
PRIOR FILING DATE: 2002-04-09
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063549
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/069873
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089905
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090691

```


PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/091992
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/095302
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/095318
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/095916
PRIOR FILING DATE: 1998-08-10
PRIOR APPLICATION NUMBER: 60/096146
PRIOR FILING DATE: 1998-08-11
PRIOR APPLICATION NUMBER: 60/096791
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: 60/097986
PRIOR FILING DATE: 1998-08-26
PRIOR APPLICATION NUMBER: 60/098544
PRIOR FILING DATE: 1998-08-31
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099803
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099811
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100038
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100848
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101741
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101786
PRIOR FILING DATE: 1998-09-25
PRIOR APPLICATION NUMBER: 60/101916
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101922
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/106178
PRIOR FILING DATE: 1998-10-28
PRIOR APPLICATION NUMBER: 60/106248
PRIOR FILING DATE: 1998-10-29
PRIOR APPLICATION NUMBER: 60/106464
PRIOR FILING DATE: 1998-10-30
PRIOR APPLICATION NUMBER: 60/106905
PRIOR FILING DATE: 1998-11-03
PRIOR APPLICATION NUMBER: 60/108787
PRIOR FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: 60/108801
PRIOR FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: 60/108849
PRIOR FILING DATE: 1998-11-18
PRIOR APPLICATION NUMBER: 60/112422
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: 60/113296
PRIOR FILING DATE: 1998-12-22

PRIOR APPLICATION NUMBER: 60/113605
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113621
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/115558
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115565
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115733
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/119549
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/123618
PRIOR FILING DATE: 1999-03-10
PRIOR APPLICATION NUMBER: 60/125259
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: 60/125775
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/126773
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: 60/127887
PRIOR FILING DATE: 1999-04-05
PRIOR APPLICATION NUMBER: 60/130232
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/131022
PRIOR FILING DATE: 1999-04-26
PRIOR APPLICATION NUMBER: 60/131270
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131291
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131445
PRIOR FILING DATE: 1999-04-28
PRIOR APPLICATION NUMBER: 60/134287
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/140650
PRIOR FILING DATE: 1999-06-22
PRIOR APPLICATION NUMBER: 60/140723
PRIOR FILING DATE: 1999-06-22
PRIOR APPLICATION NUMBER: 60/141037
PRIOR FILING DATE: 1999-06-23
PRIOR APPLICATION NUMBER: 60/144758
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/145698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 60/146222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: 60/146963
PRIOR FILING DATE: 1999-08-03
PRIOR APPLICATION NUMBER: 60/149320
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/149638
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/151733
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/164418
PRIOR FILING DATE: 1999-11-09
PRIOR APPLICATION NUMBER: 60/166361
PRIOR FILING DATE: 1999-11-16
PRIOR APPLICATION NUMBER: 60/169445
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: 60/169495
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: 60/169835

Query Match 65.5%; Score 1110.5; DB 9; length 667;
Best Local Similarity 76.6%; Pred. No. 1.6e-91;
Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYSALVLSCLITFLVLMRSIVTNR-----IGSGGSD-----GQSYNULFS 42
DB 365 CYSALVLSCLITFLVLMRSIVTNRALRGALDLSPLHRSPPHSROATFCMSFS 424
QY 43 VRLPLPLAGLLVQOIIFFLGTTALAFVLMPLVHGRNLLFFRSLSGSPFWLTALAVI 102

Db 425 AYQTAFCIGLLVQOIIFFLGTTALAFVLPVPHGRNLLFRSLSSWPFWTLALAVI 484
Qy 103 LONMAAHVPLETHDGHQOLNRRVLYAATFLLPPLNVVGAMKXXXC----- 149
Db 485 LONMAAHVPLETHDGHQOLNRRVLYAATFLLPPLNVVGAMVATWRVLSALYNAIHL 544
Qy 150 -SPSIAIRHP-----TPGYTYTRNFKIEVSQSHPAMTAFCSLLLOAOSLLPRTMAPOD 203
Db 545 GOMDSLPPRAATLDPGYTYTRNFKIEVSQSHPAMTAFCSLLLOAOSLLPRTMAPOD 604
Qy 204 SLRGEDEGMQLLOTQKDSMAKARGARPGARWGLAYTLLHNPTLOVFRKTALLGANG 263
Db 605 SLRGEDEGMQLLOTQKDSMAKARGARPGARWGLAYTLLHNPTLOVFRKTALLGANG 664
Qy 264 AQP 266
Db 665 AQP 667

RESULT 6
US-10-218-631-80
; Sequence 80, Application US/10218631
; Publication No. US20030045667A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe F.
; APPLICANT: Watanabe, Colin L.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3530PLC14
; CURRENT FILING DATE: 2002-08-12
; PRIOR FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063549
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/069873
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079728
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 246
; SEQ ID NO 80
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-218-631-80

Query Match 65.5%; Score 1110.5; DB 9; Length 667;
Best Local Similarity 76.6%; Pred. No. 1.6e-91;
Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

Qy 1 CYISALVLSCLTFLVLRSLVTR-----IGSGSGD-----GOESVNF 42
Db 365 CYISALVLSCLTFLVLRSLVTRNTRALHGAALDLSLHSPHPSROAIFCMMSFS 424
Qy 43 VPLPPLAGLLVQOIIFFLGTTALAFVLPVPHGRNLLFRSLSSWPFWTLALAVI 102
Db 425 AYQTAFCIGLLVQOIIFFLGTTALAFVLPVPHGRNLLFRSLSSWPFWTLALAVI 484
Qy 103 LONMAAHVPLETHDGHQOLNRRVLYAATFLLPPLNVVGAMKXXXC----- 149
Db 485 LONMAAHVPLETHDGHQOLNRRVLYAATFLLPPLNVVGAMVATWRVLSALYNAIHL 544
Qy 150 -SPSIAIRHP-----TPGYTYTRNFKIEVSQSHPAMTAFCSLLLOAOSLLPRTMAPOD 203
Db 545 GOMDSLPPRAATLDPGYTYTRNFKIEVSQSHPAMTAFCSLLLOAOSLLPRTMAPOD 604
Qy 204 SLRGEDEGMQLLOTQKDSMAKARGARPGARWGLAYTLLHNPTLOVFRKTALLGANG 263
Db 605 SLRGEDEGMQLLOTQKDSMAKARGARPGARWGLAYTLLHNPTLOVFRKTALLGANG 664
Qy 264 AQP 266
Db 665 AQP 667

RESULT 7
US-10-230-338-80
; Sequence 80, Application US/10230338
; Publication No. US2003004934A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe F.
; APPLICANT: Watanabe, Colin L.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3530PLC92
; CURRENT FILING DATE: 2002-08-28
; PRIOR FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: 10/119,480
; PRIOR FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063549
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/069873
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079728
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 246
; SEQ ID NO 80
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-230-338-80

Query Match 65.5%; Score 1110.5; DB 9; Length 667;
 Best Local Similarity 76.2%; Pred. No. 1.6e-91;
 Matches 221; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYSALVLSCLTFLVLMRSVLTNR-----LGGSGSGD-----GQFSWNLFS 42
 DB 365 CYSALVLSCLTFLVLMRSVLTNRVLRALHRCALDLSPLHRSPPHRSQALFCWMSFS 424
 QY 43 VPLPLPPLAGLLVQOIIFFLGTTLALAFVLMVPHGRNLLFFRSLESSWPFMTLALAVI 102
 DB 425 AVOTAFICLGLVQOIIFFLGTTLALAFVLMVPHGRNLLFFRSLESSWPFMTLALAVI 484
 QY 103 LONMAAHVFLETHDGHGHPOLTNRRVLYAATFLFPLNVLVGAMKXXC----- 149
 DB 485 LONMAAHVFLETHDGHGHPOLTNRRVLYAATFLFPLNVLVGAMVATRWVLLSALYNAIHL 544
 QY 150 -SPSIAIRHP-----TPGYTYRNFLKIEVSQSHPMATFCSLLLOQSLPRTMAAPOD 203
 DB 545 GQMDLSLPPRAATLDPGYTYRNFLKIEVSQSHPMATFCSLLLOQSLPRTMAAPOD 604
 QY 204 SLRGEDEGQMLLOTQDSNAKGPAGXGRARWGLAYTLLHNPTLOVFRKTALLGANG 263
 DB 605 SLRGEDEGQMLLOTQDSNAKGPAGXGRARWGLAYTLLHNPTLOVFRKTALLGANG 664
 QY 264 AQP 266
 DB 665 AQP 667

RESULT 8

US-09-759-056-5
 ; Sequence 5, Application US/09759056
 ; Patent No. US20020156252A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pennica, Diane
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Wood, William I.
 ; TITLE OF INVENTION: No. US20020156252A1e1 STRA6 Polypeptides
 ; FILE REFERENCE: GENENT.2827A2
 ; CURRENT APPLICATION NUMBER: US/09/759.056
 ; PRIOR FILING DATE: 2001-01-11
 ; PRIOR APPLICATION NUMBER: 60/197089
 ; PRIOR FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/175849
 ; PRIOR FILING DATE: 2000-01-13
 ; PRIOR APPLICATION NUMBER: 60/228914
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5
 ; LENGTH: 658
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-759-056-5

Query Match 65.3%; Score 1106.5; DB 9; Length 658;
 Best Local Similarity 76.2%; Pred. No. 3.5e-91;
 Matches 231; Conservative 5; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYSALVLSCLTFLVLMRSVLTNR-----LGGSGSGD-----GQFSWNLFS 42
 DB 356 CYSALVLSCLTFLVLMRSVLTNRVLRALHRCALDLSPLHRSPPHRSQALFCWMSFS 415
 QY 43 VPLPLPPLAGLLVQOIIFFLGTTLALAFVLMVPHGRNLLFFRSLESSWPFMTLALAVI 102
 DB 416 AVOTAFICLGLVQOIIFFLGTTLALAFVLMVPHGRNLLFFRSLESSWPFMTLALAVI 475
 QY 103 LONMAAHVFLETHDGHGHPOLTNRRVLYAATFLFPLNVLVGAMKXXC----- 149
 DB 476 LONMAAHVFLETHDGHGHPOLTNRRVLYAATFLFPLNVLVGAMVATRWVLLSALYNAIHL 535
 QY 150 -SPSIAIRHP-----TPGYTYRNFLKIEVSQSHPMATFCSLLLOQSLPRTMAAPOD 203

DB 536 GQMDLSLPPRAATLDPGYTYRNFLKIEVSQSHPMATFCSLLLOQSLPRTMAAPOD 595
 QY 204 SLRGEDEGQMLLOTQDSNAKGPAGXGRARWGLAYTLLHNPTLOVFRKTALLGANG 263
 DB 596 SLRGEDEGQMLLOTQDSNAKGPAGXGRARWGLAYTLLHNPTLOVFRKTALLGANG 655

QY 264 AQP 266
 DB 656 AQP 658

RESULT 9

US-09-901-812-5
 ; Sequence 5, Application US/09901812
 ; Patent No. US20020173461A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pennica, Diane
 ; APPLICANT: Polakis, Paul
 ; APPLICANT: Szeto, Wayne
 ; TITLE OF INVENTION: UPREGULATION OF TUMOR ANTIGENS TO
 ; FILE REFERENCE: GENENT.083A
 ; CURRENT APPLICATION NUMBER: US/09/901.812
 ; PRIOR FILING DATE: 2001-07-10
 ; PRIOR APPLICATION NUMBER: 60/228.914
 ; PRIOR FILING DATE: 2000-08-29
 ; PRIOR APPLICATION NUMBER: 09/759.056
 ; PRIOR FILING DATE: 2001-01-11
 ; PRIOR APPLICATION NUMBER: 60/175.849
 ; PRIOR FILING DATE: 2000-01-13
 ; PRIOR APPLICATION NUMBER: 60/197.089
 ; PRIOR FILING DATE: 2000-04-14
 ; NUMBER OF SEQ ID NOS: 28
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5
 ; LENGTH: 658
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-901-812-5

Query Match 65.3%; Score 1106.5; DB 9; Length 658;
 Best Local Similarity 76.2%; Pred. No. 3.5e-91;
 Matches 231; Conservative 5; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYSALVLSCLTFLVLMRSVLTNR-----LGGSGSGD-----GQFSWNLFS 42
 DB 356 CYSALVLSCLTFLVLMRSVLTNRVLRALHRCALDLSPLHRSPPHRSQALFCWMSFS 415
 QY 43 VPLPLPPLAGLLVQOIIFFLGTTLALAFVLMVPHGRNLLFFRSLESSWPFMTLALAVI 102
 DB 416 AVOTAFICLGLVQOIIFFLGTTLALAFVLMVPHGRNLLFFRSLESSWPFMTLALAVI 475
 QY 103 LONMAAHVFLETHDGHGHPOLTNRRVLYAATFLFPLNVLVGAMKXXC----- 149
 DB 476 LONMAAHVFLETHDGHGHPOLTNRRVLYAATFLFPLNVLVGAMVATRWVLLSALYNAIHL 535
 QY 150 -SPSIAIRHP-----TPGYTYRNFLKIEVSQSHPMATFCSLLLOQSLPRTMAAPOD 203
 DB 536 GQMDLSLPPRAATLDPGYTYRNFLKIEVSQSHPMATFCSLLLOQSLPRTMAAPOD 595
 QY 204 SLRGEDEGQMLLOTQDSNAKGPAGXGRARWGLAYTLLHNPTLOVFRKTALLGANG 263
 DB 596 SLRGEDEGQMLLOTQDSNAKGPAGXGRARWGLAYTLLHNPTLOVFRKTALLGANG 655
 QY 264 AQP 266
 DB 656 AQP 658

RESULT 10

US-09-816-653a-7
 ; Sequence 7, Application US/0981653A

```

; Publication No. US20030021788A1
; GENERAL INFORMATION:
; APPLICANT: Raselli, Luca
; APPLICANT: Pennica, Diane
; TITLE OF INVENTION: NOVEL HUMAN STRA6-LIKE PROTEIN AND NUCLEIC ACIDS ENCODING THE SAME
; FILE REFERENCE: 10716/57
; CURRENT APPLICATION NUMBER: US/09/816,653A
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/191,532
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 7
; LENGTH: 670
; TYPE: PRT
; ORGANISM: Mus musculus retinoic acid-responsive protein (Stra6) mRNA, complete cds
US-09-816-653A-7

```

```

Query Match          49.1%; Score 832.5; DB 9; Length 670;
Best Local Similarity 57.7%; Pred. No. 1.2e-66;
Matches 176; Conservative 28; Mismatches 62; Indels 39; Gaps 4;

```

```

QY 1 CYISALVISCCLTFLVLMSPVLTNR-----LQSGGSGD-----GQPSWNLFS 42
DB 366 CYISALVISCSTFLLRLSLRTHRANLQALHRGALDLPPLQSIHPSRQALVSWMSFC 425
QY 43 VPLPLPLAGLVQOIIIFLGTTLAFLVLMPLVHGRNLLFRSLSSMPFWLTLALVI 102
DB 426 AVQTAFSCGLLVQVIFFLGTSLAFLVPLVHGRNLLRLSLSTWPFVLTALVI 485
QY 103 LQNMMAHWFLETHDGHQPOLTNRRVLYAATFLLFPLNLVGMXXXXCSPSIAIRHPT--- 159
DB 486 LQNIANMIPLRTHHGYBELTNRRLCVATFLLFPIINMLVGMVWVRLISSLVNTHL 545
QY 160 -----PGVYTRNFKLIEVSQSHPMATFCSLLAQSLLPRTMAAD 203
DB 546 GMDLSLPORASLDPGYHTQNTLRLEASQSHPGVIAFCALLHAASPOPRPLAPOD 605
QY 204 SLRPEDEGMQLQTKDSMAKGARPGAKRGARWGLAYTLLHNPFLQVFRKTLALG--A 261
DB 606 SLRPAEEBGMQLQTKDLMAGKAGKSGSRAKGLAYTLLHNSLQAFKRALTSKA 665
QY 262 NGAOP 266
DB 666 NGTOP 670

```

```

RESULT 11
US-09-925-299-796
; Sequence 796, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 796
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-796

```

```

Query Match          28.6%; Score 485; DB 9; Length 178;
Best Local Similarity 88.2%; Pred. No. 3.3e-36;
Matches 97; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

```

QY 36 FSNWLFSPVPLPLAGLVQOIIIFLGTTLAFLVLMPLVHGRNLLFRSLSSMPFWL 95
DB 27 FSNWLFSPVPLPLAGLVQOIIIFLGTTLAFLVLMPLVHGRNLLFRSLSSMPFWL 86
QY 96 TLALAVILQNMMAHWFLETHDGHQPOLTNRRVLYAATFLLFPLNLVGM 145
DB 87 TLALAVILQNMMAHWFLETHDGHQPOLTNRRVLYAATFLLFPLNLVGM 136

```

```

RESULT 12
US-09-925-299-796
; Sequence 796, Application US/09925299
; Patent No. US2002005627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 796
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-796

```

```

Query Match          28.6%; Score 485; DB 10; Length 178;
Best Local Similarity 88.2%; Pred. No. 3.3e-36;
Matches 97; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

```

QY 36 FSNWLFSPVPLPLAGLVQOIIIFLGTTLAFLVLMPLVHGRNLLFRSLSSMPFWL 95
DB 27 FSNWLFSPVPLPLAGLVQOIIIFLGTTLAFLVLMPLVHGRNLLFRSLSSMPFWL 86
QY 96 TLALAVILQNMMAHWFLETHDGHQPOLTNRRVLYAATFLLFPLNLVGM 145
DB 87 TLALAVILQNMMAHWFLETHDGHQPOLTNRRVLYAATFLLFPLNLVGM 136

```

```

RESULT 13
US-09-826-508-30
; Sequence 30, Application US/09826508
; Patent No. US20010025099A1
; GENERAL INFORMATION:
; APPLICANT: Nabli Elshourbagy
; APPLICANT: Lisa Vawter
; TITLE OF INVENTION: G-Protein-Coupled Receptor Polypeptides
; TITLE OF INVENTION: and Polynucleotides
; FILE REFERENCE: GP-7074USB
; CURRENT APPLICATION NUMBER: US/09/826,508
; CURRENT FILING DATE: 2001-04-05
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 30
; LENGTH: 403
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-826-508-30

```

```

Query Match          5.8%; Score 97.5; DB 10; Length 403;
Best Local Similarity 19.9%; Pred. No. 0.46;
Matches 69; Conservative 30; Mismatches 107; Indels 14; Gaps 14;

```

```

QY 5 ALVISCCLTFLVLMSPVLTNR-----LQSGGSGD-----GQPSWNLFSVPLPLPLAGLVQOII----- 59
DB 134 ALVISCCLSGAMVRLVNH-----GTGPRAGQOLVGLALCL-----MLVQYIIIVENVL 181
QY 60 -----FFLGTTALAFVLMPLV-----VLHGRNLLFRS 86

```

```
Db 182 VLTVLRTDTPACAYEPMDFWALIVDWLVLTGLLFTLCGKFKRWKLGAFLLITAF 241
Qy 87 LE-SSWPFMLTL-----ALAVILQNMMAHWVLETHDGHQPOLNRRVL 128
Db 242 LSVLIWAMMTMYLFGNVKLOOGDAMNDPTLAILT--AASGVVFIH-AIPEI----- 292
Qy 129 YAATFLLPLNLVVGAMXXXCPSIAIRHPTPGYTYRNFLKIEVSOSHPRMTAFCSLL 188
Db 293 -----HCTLLPALQENTPNYF-----DTSQPRMRETAFA----- 320
Qy 189 QAOSLLPRTMAAPODSLRPGEEDEGMQLLOTKDSMAKARGARARGLAYTLIHN 248
Db 321 EEDVOLPRAYM-----ENKAFSMDENHNAALRTAGFPNGSLGK----- 357
Qy 249 TLQVFRKTALLGANGAOPCSSLPGPSITPRMO-----PAGPDDHXG 291
Db 358 -----RPSGSLGKRPSAPFRSNVYQPTEMAVALNGCTIPTAPPSHTG 399
```

```
RESULT 14
US-09-895-686-5
; Sequence 5, Application US/09895686
; Patent No. US20020106655A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Tang, Y. Tom
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: HUMAN GPCR PROTEINS
; FILE REFERENCE: PC-0044 CIP
; CURRENT APPLICATION NUMBER: US/09/895,686
; NUMBER OF SEQ. ID NOS: 74
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incycle ID No. US20020106655A1 2705201CD1
US-09-895-686-5
```

```
Query Match 5.8%; Score 97.5; DB 10; Length 403;
Best Local Similarity 19.9%; Pred. No. 0.46; Indels 141; Gaps 14;
Matches 69; Conservative 30; Mismatches 107;
```

```
Qy 5 ALVLSCLLTFLVLMRSIVTNRHLSGSGSDGQFSWNLFVPLPLPPLAGLIVQOII----- 59
Db 134 ALCPSCILSQAMRVRLVRH-----GTGPAGQVLGLALCL-----MLVQVIAVEM 181
Qy 60 -----FLIGTTALAFVLM-----VLHGRNLLFFRS 86
Db 182 VLTVLRTDTPACAYEPMDFWALIVDWLVLTGLLFTLCGKFKRWKLGAFLLITAF 241
Qy 87 LE-SSWPFMLTL-----ALAVILQNMMAHWVLETHDGHQPOLNRRVL 128
Db 242 LSVLIWAMMTMYLFGNVKLOOGDAMNDPTLAILT--AASGVVFIH-AIPEI----- 292
Qy 129 YAATFLLPLNLVVGAMXXXCPSIAIRHPTPGYTYRNFLKIEVSOSHPRMTAFCSLL 188
Db 293 -----HCTLLPALQENTPNYF-----DTSQPRMRETAFA----- 320
Qy 189 QAOSLLPRTMAAPODSLRPGEEDEGMQLLOTKDSMAKARGARARGLAYTLIHN 248
Db 321 EEDVOLPRAYM-----ENKAFSMDENHNAALRTAGFPNGSLGK----- 357
Qy 249 TLQVFRKTALLGANGAOPCSSLPGPSITPRMO-----PAGPDDHXG 291
Db 358 -----RPSGSLGKRPSAPFRSNVYQPTEMAVALNGCTIPTAPPSHTG 399
```

```
RESULT 15
US-09-826-508-32
; Sequence 32, Application US/09826508
; Patent No. US20010025099A1
; GENERAL INFORMATION:
; APPLICANT: Nabil Elshourbasy
; APPLICANT: Lisa Vawter
; TITLE OF INVENTION: G Protein-Coupled Receptor Polypeptides
; FILE REFERENCE: GP-70744USB
; CURRENT APPLICATION NUMBER: US/09/826,508
; NUMBER OF SEQ. ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 427
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-826-508-32
```

```
Query Match 5.8%; Score 97.5; DB 10; Length 427;
Best Local Similarity 19.9%; Pred. No. 0.49; Indels 141; Gaps 14;
Matches 69; Conservative 30; Mismatches 107;
```

```
Qy 5 ALVLSCLLTFLVLMRSIVTNRHLSGSGSDGQFSWNLFVPLPLPPLAGLIVQOII----- 59
Db 158 ALCPSCILSQAMRVRLVRH-----GTGPAGQVLGLALCL-----MLVQVIAVEM 205
Qy 60 -----FLIGTTALAFVLM-----VLHGRNLLFFRS 86
Db 206 VLTVLRTDTPACAYEPMDFWALIVDWLVLTGLLFTLCGKFKRWKLGAFLLITAF 265
Qy 87 LE-SSWPFMLTL-----ALAVILQNMMAHWVLETHDGHQPOLNRRVL 128
Db 266 LSVLIWAMMTMYLFGNVKLOOGDAMNDPTLAILT--AASGVVFIH-AIPEI----- 316
Qy 129 YAATFLLPLNLVVGAMXXXCPSIAIRHPTPGYTYRNFLKIEVSOSHPRMTAFCSLL 188
Db 317 -----HCTLLPALQENTPNYF-----DTSQPRMRETAFA----- 344
Qy 189 QAOSLLPRTMAAPODSLRPGEEDEGMQLLOTKDSMAKARGARARGLAYTLIHN 248
Db 345 EEDVOLPRAYM-----ENKAFSMDENHNAALRTAGFPNGSLGK----- 381
Qy 249 TLQVFRKTALLGANGAOPCSSLPGPSITPRMO-----PAGPDDHXG 291
Db 382 -----RPSGSLGKRPSAPFRSNVYQPTEMAVALNGCTIPTAPPSHTG 423
```

```
Search completed: March 13, 2003, 11:37:59
Job time : 15.645 secs
```


Gencore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:27:42 ; Search time 16.7462 Seconds

(without alignments)
571.023 Million cell updates/sec

Title: US-09-816-653a-4

Perfect score: 1695
Sequence: 1 CTHSALVLSCLTFLVLMRS.....APPTWMSQGFRRKKVG 325

Scoring table: BLOSUMP62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep: *
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep: *
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep: *
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep: *
6: /cgn2_6/ptodata/2/iaa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93	5.5	2016	US-09-634-920-4	Sequence 4, Appli
2	87	5.1	618	US-08-866-381A-6	Sequence 6, Appli
3	87	5.1	621	US-08-208-887A-49	Sequence 49, Appli
4	87	5.1	621	US-09-280-598-18	Sequence 18, Appli
5	87	5.1	621	US-08-945-771-4	Sequence 4, Appli
6	83	4.9	539	US-09-036-987A-11	Sequence 11, Appli
7	83	4.9	539	US-09-370-700-11	Sequence 11, Appli
8	83	4.9	1495	US-08-462-467B-12	Sequence 12, Appli
9	82.5	4.9	1024	US-09-562-737-84	Sequence 84, Appli
10	82.5	4.9	2972	US-08-469-260A-387	Sequence 387, App
11	82.5	4.9	5087	US-09-144-085-1	Sequence 1, Appli
12	81.5	4.8	376	US-08-875-982-2	Sequence 2, Appli
13	81.5	4.8	6095	US-09-144-085-2	Sequence 2, Appli
14	81	4.8	410	US-09-083-521-2	Sequence 2, Appli
15	81	4.8	419	US-08-439-131A-3	Sequence 3, Appli
16	81	4.8	419	US-08-440-674-2	Sequence 2, Appli
17	80	4.7	859	US-08-395-580-2	Sequence 2, Appli
18	80	4.7	859	PCT-US95-02792-2	Sequence 2, Appli
19	80	4.7	2618	US-09-413-814-28	Sequence 28, Appli
20	79.5	4.7	2588	US-08-936-135-2	Sequence 2, Appli
21	79	4.7	235	US-07-940-605A-12	Sequence 12, Appli
22	79	4.7	235	US-08-690-096-12	Sequence 12, Appli
23	79	4.7	1306	US-08-999-774A-13	Sequence 13, Appli
24	79	4.7	1385	US-08-687-399-7	Sequence 7, Appli
25	77.5	4.6	376	US-08-751-512-8	Sequence 8, Appli
26	77.5	4.6	869	US-08-188-582-32	Sequence 32, Appli
27	77.5	4.6	869	US-08-646-715-32	Sequence 32, Appli

28	77	4.5	448	1	US-08-207-904-12	Sequence 2, Appli
29	77	4.5	448	1	US-08-207-904-17	Sequence 17, Appli
30	77	4.5	1017	4	US-09-600-776-6	Sequence 6, Appli
31	76.5	4.5	355	1	US-07-759-568-1	Sequence 1, Appli
32	76.5	4.5	355	1	US-08-450-393A-8	Sequence 8, Appli
33	76.5	4.5	355	2	US-08-390-000A-5	Sequence 5, Appli
34	76.5	4.5	355	4	US-08-446-669-8	Sequence 8, Appli
35	76.5	4.5	355	5	PCT-US95-00476-8	Sequence 8, Appli
36	76.5	4.5	360	1	US-08-202-056-7	Sequence 7, Appli
37	76	4.5	194	4	US-08-959-004-9	Sequence 9, Appli
38	76	4.5	195	4	US-09-042-071-41	Sequence 41, Appli
39	76	4.5	350	1	US-08-458-067-2	Sequence 2, Appli
40	76	4.5	350	5	PCT-US96-07795-2	Sequence 2, Appli
41	76	4.5	350	5	PCT-US96-07796-2	Sequence 2, Appli
42	76	4.5	635	1	US-07-832-855-2	Sequence 2, Appli
43	76	4.5	635	4	US-08-176-320-2	Sequence 2, Appli
44	76	4.5	724	4	US-08-914-999-4	Sequence 4, Appli
45	75.5	4.5	197	3	US-09-010-809-5	Sequence 5, Appli

ALIGNMENTS

```

RESULT 1
US-09-634-920-4
; Sequence 4, Application US/09634920
; Patent No. 6342357
; GENERAL INFORMATION:
; APPLICANT: Splawski, Igor
; TITLE OF INVENTION: ALTERATIONS IN THE LONG QT SYNDROME GENES KVLQT1 AND
; FILE REFERENCE: 2323-155
; CURRENT APPLICATION NUMBER: US/09/634,920
; PRIOR FILING DATE: 2000-08-09
; PRIOR APPLICATION NUMBER: 60/190,057
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/147,488
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2016
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-634-920-4

Query Match 5.5%; Score 93; DB 4; Length 2016;
Best Local Similarity 23.8%; Pred. No. 0.85;
Matches 67; Conservative 31; Mismatches 113; Indels 70; Gaps 13;

QY 35 QPSNLSVPLPLPLAGLVQIIIFLGTTALAFLVLMVLRNLLFPRSLSSWPFW 94
DB 778 QCGNITD-----SIIYIISLMELGSRMSNL--SVLSFRLLRFKLKXSPPTL 825
QY 95 -----LTALAVALIQNMAAMV-----FLETHDGHPOUTNR---RVLY 129
DB 826 NTLKIITGNSVGLAGNLTIVATIVFIFAVVGMLPFGKNVSELRDSGLPRHNMDF 885
QY 130 AATLPLPLNVLCAMKXKSPSIAIRHPPGYTYNPLKIEVSQSHPAMTAFCSILLQ 189
DB 886 HAFIIF--RILCEWETETWDCMEVSGQS--LCLLVFLVWVIGMLVVLNLFALLS 940
QY 190 AQSLLPRTMAPOSLRGERDEGMQLQTKDSMAKGRPGAKRGARWGLATYLL--HNP 248
DB 941 SFS-----ANLTAPEDEKEMNNLQ--ALARIQGLRFVYKRTTWDFCCGLRHRP 989
QY 249 TLQVFRKATLIGANGAOP--CSLPGSPSPSITPAMQAPPPD 288
DB 990 -----QKPALAAQGLPSCATATPSPP-----PPE 1015

RESULT 2

```

US-08-866-381A-6
; Sequence 6, Application US/08866381A
; Patent No. 6045797
; GENERAL INFORMATION:
; APPLICANT: Ben Lewis Margolis
; APPLICANT: Joseph Schlesinger
; TITLE OF INVENTION: METHODS FOR TREATMENT OR DIAGNOSIS
; TITLE OF INVENTION: OF DISEASES OR CONDITIONS ASSOCIATED
; TITLE OF INVENTION: WITH A BLM DOMAIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Paeleco for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/866.381A
; FILING DATE: May 30, 1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/212,234
; FILING DATE: March 14, 1994
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maiburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 226/043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 618 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; OTHER INFORMATION: GRB-10
; US-08-866-381A-6
Query Match 5.1%; Score 87; DB 3; Length 618;
Best Local Similarity 21.7%; Pred. No. 0.75;
Matches 38; Conservative 28; Mismatches 59; Indels 50; Gaps 8;
QY 172 EVSOSHPAMTAFCSL-----ILQ-----AQSLLPRMAPODSLRCGEDEG 213
Db 1 DINSVESLNSACNMOSTDTAPLEDDGQHASNOGAASSSRGQOPASPRQKRSQPVHI 60
QY 214 MQLLOTDSMAKAGAPGAXRGARWGLAYTLHNPTLQVFRKTAALLGANGAOPCSSLPGS 273
Db 61 LRLQEBDQQLRTA-----SLPAIPNPFPE-----LTGA-----APGS 93
QY 274 PPSITPAM---QPAQPP-DHXGXVEVCLHWPXXGSAF-PTWLWESQGGFMRKKL 323
Db 94 PPSVAPSSLPPPPSQPAKHGCRCEKWIPEGNTRGNKRIKWQFPFGFQLSKL 148
RESULT 3
US-08-208-887A-49
; Sequence 49, Application US/08208887A
; Patent No. 5677421

GENERAL INFORMATION:
; APPLICANT: Schlesinger, Joseph
; APPLICANT: Skolnick, Edward Y.
; APPLICANT: Margolis, Benjamin L.
; TITLE OF INVENTION: NOVEL EXPRESSION CLONING METHOD FOR
; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
; TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: 10036-2711
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/208,887A
; FILING DATE: 11-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7693-063
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 621 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-208-887A-49
Query Match 5.1%; Score 87; DB 1; Length 621;
Best Local Similarity 21.7%; Pred. No. 0.75;
Matches 38; Conservative 28; Mismatches 59; Indels 50; Gaps 8;
QY 172 EVSOSHPAMTAFCSL-----ILQ-----AQSLLPRMAPODSLRCGEDEG 213
Db 4 DINSVESLNSACNMOSTDTAPLEDDGQHASNOGAASSSRGQOPASPRQKRSQPVHI 63
QY 214 MQLLOTDSMAKAGAPGAXRGARWGLAYTLHNPTLQVFRKTAALLGANGAOPCSSLPGS 273
Db 64 LRLQEBDQQLRTA-----SLPAIPNPFPE-----LTGA-----APGS 96
QY 274 PPSITPAM---QPAQPP-DHXGXVEVCLHWPXXGSAF-PTWLWESQGGFMRKKL 323
Db 97 PPSVAPSSLPPPPSQPAKHGCRCEKWIPEGNTRGNKRIKWQFPFGFQLSKL 151
RESULT 4
US-09-280-598-18
; Sequence 18, Application US/09280598
; Patent No. 6391584
; GENERAL INFORMATION:
; APPLICANT: Schlesinger, Joseph
; APPLICANT: Skolnick, Edward Y.
; APPLICANT: Margolis, Benjamin L.
; TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR
; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE
; TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas


```

CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/280,598
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/252,820
FILING DATE: 02-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Cortuzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7683-067
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 621 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-09-280-598-18

Query Match          5.1%; Score 87; DB 4; Length 621;
Best Local Similarity 21.7%; Pred. No. 0.75;
Matches 38; Conservative 28; Mismatches 59; Indels 50; Gaps 8;

QY 172 EVSOSHPMATFCSL-----LLQ-----AQSLLPRTMAAPODSLRPEDEEG 213
DB 4 DINSSVESLNSACNMQSDTPTAPLLEDQGHASNGQAASSSSRGQOPASPRQMQRSGPVHI 63
QY 214 MQLTQKSMAGARPGAXRGRARWGLAVTLHNPTLOVFRKTLALLGANGAPCCSLPGS 273
DB 64 LRRLEEDQQLRTA-----SLPAIPNPFPE-----LTGA-----APGS 96
QY 274 PPSITPAM---QPAQPP-DHXGXVEVCILHMEPPXXGSAP-PTWLMSQCGFWRKLU 323
DB 97 PPSVAPSSLPPEPSPAPKAGCRCEKWIPEGENTRGKRIWRQFPFGFOLSKL 151

RESULT 5
US-08-945-771-4
Sequence 4, Application US/08945771
Patent No. 6465623
GENERAL INFORMATION:
APPLICANT: Daly, Roger J
APPLICANT: Sutherland, Robert L
TITLE OF INVENTION: GDU, A novel signalling protein
FILE REFERENCE: 273402001700
CURRENT APPLICATION NUMBER: US/08/945,771
CURRENT FILING DATE: 1998-04-22
EARLIER APPLICATION NUMBER: PCT/US96/00258
EARLIER FILING DATE: 1996-MAY-02
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patentin Ver. 2.1
SEO ID NO 4
LENGTH: 621
TYPE: PRT
ORGANISM: Mus musculus
US-08-945-771-4

Query Match          5.1%; Score 87; DB 4; Length 621;
Best Local Similarity 21.7%; Pred. No. 0.75;
Matches 38; Conservative 28; Mismatches 59; Indels 50; Gaps 8;
```

```

QY 172 EVSOSHPMATFCSL-----LLQ-----AQSLLPRTMAAPODSLRPEDEEG 213
DB 4 DINSSVESLNSACNMQSDTPTAPLLEDQGHASNGQAASSSSRGQOPASPRQMQRSGPVHI 63
QY 214 MQLTQKSMAGARPGAXRGRARWGLAVTLHNPTLOVFRKTLALLGANGAPCCSLPGS 273
DB 64 LRRLEEDQQLRTA-----SLPAIPNPFPE-----LTGA-----APGS 96
QY 274 PPSITPAM---QPAQPP-DHXGXVEVCILHMEPPXXGSAP-PTWLMSQCGFWRKLU 323
DB 97 PPSVAPSSLPPEPSPAPKAGCRCEKWIPEGENTRGKRIWRQFPFGFOLSKL 151

RESULT 6
US-09-036-987A-11
Sequence 11, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patil J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESS: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 539 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-036-987A-11

Query Match          4.9%; Score 83; DB 4; Length 539;
Best Local Similarity 19.9%; Pred. No. 1.7;
Matches 62; Conservative 38; Mismatches 122; Indels 90; Gaps 16;
```

Qy	137	PLNVLVGANXXXXCS	SIARHPTTPTTYRNRFLKEIVSQSHRPMATPFCSLLOAQSLPR	166
Db	286	PESTOLGLFS----	TLCAHRQAYV-----LVNHLGDTP--NAERTTAEHLSSAINAQ	334
Qy	197	TMAAPODSLRPEEBDEG	QOLLQTKDSMKAGARPGAXRGRARNGLAYTLTLLNHPVL--QYFRK	255
Db	335	VGVTVAEGR---	ETPLMLRSTQVNAIALGEGEPBCMQRTKYKAAAYLRTGLSEAOQATAYRR	392
Qy	256	TALLGANGAQCPCSSL----	PCSPSTPRAMQAPRPDHXKXVEVCLHW--EPXXGSAPR	309
Db	393	LTVGVYD--NPA	ALLLLGLGGMANAVAPATALLAQDSVLKALFVNNMBPBADEHNLT	450
Qy	310	WLMESQSGCFMRK	321	
Db	451	WI----	RGFYRE	458

```

RESULT 7
US-09-370-700-11
; Sequence 11, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Tredeaway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clave
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 539
; TYPE: PRF
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-11

```

Query Match 4.9%; Score 83; DB 4; Length 539;
Best Local Similarity 19.9%; Pred. NO. 1.7;
Matches 62; Conservative 38; Mismatches 122; Indels 90; Gaps 16

QY	30	GGGDGQPS-----W-----NLFSVP-----LPLRLDGLVQQLIFFLGLTALAFVLMFVL	76
Db	217	GGGGNFGIVTRFMILRTDVPVSTDAELLPRPVALV-----	25
QY	77	HORNLLFRSLSSPFMLTALAVILIONMAHWLETHDGHQLTNRVLVAATLLF	136
Db	255	-----RSFHPMHELTQSFALLQNF-GNW--YEQHA-----	289
QY	137	PLNLVVGAMXXXCPSIAIRPTPGYTYRNFLEKVEGSHAMTAPQSLLOQSLPR	196
Db	286	PESTQGLGFS-----TLVCARQAGYV-----LVNHLDGTP--NAERTLAHEHLSALNAQ	334
QY	197	TMAPQDRLRGEEDEGQQLIQTKDSNAKGRPGAKRGARAGLAYTLTLNPTL-QVFR	255
Db	335	VGVTPEAEELR--ETLPMILRSTQVGAILEGGEPEQMORTKYAAALRTGLSEAGLATVYRR	352
QY	256	TALLGANGAQCSSL-----PGSPPSITPAMQAGPPDHGXVVEVCLHW--EPKXGSA-PT	309
Db	393	LTVGYD--NPAALLLLLLGYGGMANAVAPSTALAQRDVYLKALFTVTNNWPEADERHLLT	450
QY	310	MLMESQCGFMARK 321	
Db	451	WI-----RGFYRE 458	

RESULT 8

```

US-08-462-467B-12
; Sequence 12, Application US/08462467B
; Patent No. 6210899
; GENERAL INFORMATION:
; APPLICANT: Rosenbaum, Jan S
; TITLE OF INVENTION: The Use of a BMP Protein Receptor
; TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
; TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 11810 East Miami River Road
; CITY: Rose
; STATE: OH
; COUNTRY: USA
; ZIP: 45061
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,467B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hersko, Bart S.
; REGISTRATION NUMBER: 32,572
; REFERENCE/DOCKET NUMBER: 5474R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-0633
; TELEFAX: (513) 627-0260
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1495 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-462-467B-12

```

Query Match	4.9%	Score 83;	DB 4;	Length 1495;
Best Local Similarity	24.4%	Pred. No. 6.6;		
Matches 74;	Conservative 32;	Mismatches 117;	Indels 80;	Gaps 13

[illegible]

RESULT 9
US-09-562-737-84

; Sequence 84, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562.737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 84
; LENGTH: 1024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-84

Query Match 4.9%; Score 82.5; DB 4; Length 1024;
Best Local Similarity 19.8%; Pred. No. 4.5;
Matches 51; Conservative 36; Mismatches 85; Indels 85; Gaps 10;

QY 35 QFSWNLSPVLPPLAGLVQIIFLGTALAFVLMPLVHGNNLFFRSLSSESPFW 94
DB 772 QENNIFD-----GIIVQLSIMEIG--LAVEGILSVLRSSRLRVFKLATSWP-T 818
QY 95 LTLAVLWMAAHVWLEHNDHQPQTLNRVLYATFLLPLNLV--VGAMXXXXCSPS 152
DB 819 LNMILVIGNSVGLW-----NLTVLATITFFFAVVGMAFGKSYECVDK 865
QY 153 IAIRPTP-----GYTYR-----NFLKIEVSQSHP 178
DB 866 INVQCKLEPMHNDPEHSGILVFRVLCGHWJETWMDCMIVAGQTCILKFMVWVIGMMV 925
QY 179 AMTAFCSLLAQSLPRTMAAPQDSLRGEBDEGMQLQTKDSMAKARPGAKGRARW 238
DB 926 VLNFLFLNLSSFS-----SDNLQATDDDNEMNRLQ-----IAVGRMOS 964
QY 239 GLAVTLNHPFLQVPRK 255
DB 965 GIDF--VQNKITECPRK 979

RESULT 10
US-08-469-260A-387
; Sequence 387, Application US/08469260A
; Patent No. 6451578
; GENERAL INFORMATION:
; APPLICANT: JOHN N. SIMONS
; APPLICANT: TAMI J. PILOT-MATIAS
; APPLICANT: GEORGE J. DAWSON
; APPLICANT: GEORGE G. SCHLAUDER
; APPLICANT: THOMAS M. DESAI
; APPLICANT: TURESH M. LEARY
; APPLICANT: ANTHONY SCOTT MUEHROFF
; APPLICANT: JAMES C. BRKER
; APPLICANT: SHERI L. BUIJK
; APPLICANT: ISA K. MUSHAMAR
; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
; TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
; NUMBER OF SEQUENCES: 716
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,260A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,550
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 387:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2972 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-469-260A-387

Query Match 4.9%; Score 82.5; DB 4; Length 2972;
Best Local Similarity 24.8%; Pred. No. 19;
Matches 50; Conservative 14; Mismatches 75; Indels 63; Gaps 9;

QY 132 TELLPPLNLVGMANXXXXSPSIAIRHPPTGYTYRNLFLKIEV-SQSHPRAMTAFCSLLQA 190
DB 2800 TILMTP-----RHPVRYVLIPIHVLIMVRSQSTPDELVMCCVQSNH 2841
QY 191 QS-----LPRMTAA-----PODSLRPCEE-----DEGMQLQTKDSMAKARPGA 231
DB 2842 YSPFLRLPRVLVSLHGWCLQVTTDSIKTMEAGSXLRDLSLWRRRRAGVTRRL 2901
QY 232 XRGARWG-LAYTLNHPFLQVFRKTLALGANAPCSSLPSPSITPRAMQACPRPHX 290
DB 2902 LRGSKEWGHLLARALLMXEXLK-----EXPPIINSLPG-----POLATPYEHN 2943
QY 291 GXVEYCLHWPBXGAPPTWLM 312
DB 2944 EEWLISIK-----SRPWIRM 2959

RESULT 11
US-09-144-085-1
; Sequence 1, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Jullen, Bryan
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; FILE REFERENCE: 30062-20020.20
; CURRENT APPLICATION NUMBER: US/09/144,085
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5087
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-144-085-1
Query Match 4.9%; Score 82.5; DB 4; Length 5087;
Best Local Similarity 29.2%; Pred. No. 39;
Matches 33; Conservative 12; Mismatches 47; Indels 21; Gaps 5;

Oy 143 GAMXXCSPSIAIRHPTGYTYRNFLEKIEVSQSHPA--MTAFCSLLQAO--SLPRTMA 199
 Db 224 GALAVFPGRAKRP-----LYLSSKSNLGHGTGPAGVGVLMVSMQHEVLPRTILH 288
 Oy 200 APQDSLPRGEDEGMQLQTKDSMAKARPAXRGRAR-----WGLAYTLILH 246
 Db 289 AEQPSPHIGWEGSGLSLQE-----ARPMRNGRARRAGVSGISGTNAH 334

RESULT 12

US-08-875-972-2
 ; Sequence 2, Application US/08875972
 ; Patent No. 5985564
 ; GENERAL INFORMATION:
 ; APPLICANT: Huntington Porter and Jinhue Li
 ; TITLE OF INVENTION: ASSAY FOR IDENTIFYING GENES CAUSING
 ; NUMBER OF SEQUENCES: 29
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
 ; STREET: Two Millicia Drive
 ; CITY: Lexington
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02173-4799
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/875,972
 ; FILING DATE: 08-AUG-97
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/002,448
 ; FILING DATE: 16-AUG-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Granahan Esq., Patricia
 ; REGISTRATION NUMBER: 32,227
 ; TELEPHONE: (781) 861-9540
 ; TELEFAX: (781) 861-6240
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (781) 861-9540
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 376 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-875-972-2

Query Match 4.8%; Score 81.5; DB 2; Length 376;

Best Local Similarity 20.1%; Pred. No. 1.5;
Matches 65; Conservative 35; Mismatches 114; Indels 109; Gaps 13;

Oy 1 CY-----ISALVLSCLTFVLVMSRVTHRLGS-----GSGDGGFQSNWL 40
 Db 98 CYKTFHGMILMSLMLLFETIYIGEVIKTYNVAMDYTLTLTVNFGAVHGVIHWK- 156
 Oy 41 FSVPLPLPPLAGLIVQOILFFIGTTALAFVMPVLHGRLNLFRRSLSSSPFWTLA-- 98
 Db 157 -----GPIVLEQAYLMISA-----LMP-----LMFIKYPPMGSAMVILAPI 193
 Oy 99 ---LAVILQNNAAHWFLETHDGHFQLTNRNVLYAA---TFLLFPLN-VLVGAMXXXC 149
 Db 194 SYVDLVATVLCSTGPLRLMIVETAQERNETIFSPLIYSSPMVMTVMMSKLDPSQCALQJFY 253
 Oy 150 SPSIAIRHPTGYTYRNFLEKIEVSQSHPA--MTAFCSLLQAO--SLPRTMA 209
 Db 254 DPEMEDSYDSFGESPYPVEF-----PPLAGVGEELIEEBE 289

Oy 210 EDEG--MQLQTKDSMAKARPAXRGRAR-----GLAYTLINPTLQVERK 255
 Db 290 ESQGVKVLGTSISTVVPGGQAPPTSGDWITTLACFVALILGICLTL---LLAVFKX 346
 Oy 256 TALGANGAQPCSSLPSSPSIT 278
 Db 347 -----ALPALPISIT 356

RESULT 13

US-09-144-085-2
 ; Sequence 2, Application US/09144085
 ; Patent No. 6280999
 ; GENERAL INFORMATION:
 ; APPLICANT: Gustafsson, Claes
 ; APPLICANT: Ashley, Gary
 ; APPLICANT: Ziermann, Rainer
 ; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
 ; FILE REFERENCE: 30062-20020, 20
 ; CURRENT APPLICATION NUMBER: US/09/144,085
 ; EARLIER FILING DATE: 1998-08-31
 ; EARLIER APPLICATION NUMBER: 09/010,809
 ; EARLIER FILING DATE: 1998-01-22
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 6095
 ; TYPE: PRT
 ; ORGANISM: Sorangium cellulosum
 ; US-09-144-085-2

Query Match 4.8%; Score 81.5; DB 4; Length 6095;

Best Local Similarity 29.2%; Pred. No. 64;
Matches 33; Conservative 12; Mismatches 47; Indels 21; Gaps 5;

Oy 143 GAMXXCSPSIAIRHPTGYTYRNFLEKIEVSQSHPA--MTAFCSLLQAO--SLPRTMA 199
 Db 352 GALVVFPGRAKRP-----LYLSSKSNLGHGTGPAGVGVLMVSMQHEVLPRTILH 406
 Oy 200 APQDSLPRGEDEGMQLQTKDSMAKARPAXRGRAR-----WGLAYTLILH 246
 Db 407 AEQPSPHIGWEGSGLSLQE-----ARPMRNGRARRAGVSGISGTNAH 452

RESULT 14

US-09-083-521-2
 ; Sequence 2, Application US/09083521
 ; Patent No. 6048970
 ; GENERAL INFORMATION:
 ; APPLICANT: Lal, Preeti
 ; APPLICANT: Guegler, Karl J.
 ; APPLICANT: Corley, Neil C.
 ; TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 ; STREET: 3174 PORTER DRIVE
 ; CITY: PALO ALTO
 ; STATE: CALIFORNIA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/083,521
 ; FILING DATE: Herewith
 ; CLASSIFICATION:

```
ATTORNEY/AGENT INFORMATION:
NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0527 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTTUT03
CLONE: 1999442
US-09-083-521-2

Query Match      4.8%; Score 81; DB 3; Length 410;
Best Local Similarity 20.4%; Pred. No. 1.9;
Matches 73; Conservative 47; Mismatches 123; Indels 114; Gaps 17;

QY 1 CYSALVLSCLTFPLVLMRSIVTRHSGSGDGFMSNLFVPLPLPLAGL--LVQOI 58
DB 34 CDOPGIVFCTMDYDVOFC-----DPLGSLMSVWVTYAMRLQPVVXQV 79
QY 59 IFFLGTALAFVLM-----VLHGRNLPFRSLSSW-----PFW----- 94
DB 80 LYLGAALMSAQLDHRGLMNLGSPSLFALGLATMTVSVRRHRCYPTWRKMLFYL 139
QY 95 ----LTLALAVILONMAAHVFLETHDGH---POLTRNRVLYAATFLIFPL----NVLV 142
DB 140 CPGSLIGSASVLT-----YAFVETRDNYFYIHSIMHMLIGSVGFLPLPPAKTDHGVPS 193
QY 143 GAMXXCSPSIAL-----RHPTPGYVT-----YRNLKIVSQSHR 178
DB 194 GARRGGCYOICIEQEPGPRGRGHCOQHLQALGALALRGECFLEFLGWSP 253
QY 179 AMTAFCSILLQASILPRTMAAPODSLRP---GEDEGMQLLOTKDSMAK-GARPGAXR 233
DB 254 -LRRRQAVFLEDMSFRTQSSRD-LEFPFGHELEPGLESFCIMESFLRTGAVAGRES 311
QY 234 GRAR-----WGLAYTILHNPLOVFRKTXALLGANGAOPCSSLPG--SPSITPAMQ 282
DB 312 LRTESILQWSSISW-----DAEPSQDMSPFGROSPVASTASFO 351

RESULT 15
US-08-439-131A-3
Sequence 3, Application US/08439131A
Patent No. 5512472
GENERAL INFORMATION:
APPLICANT: Lai, Margaret H. K.
APPLICANT: Bard, Martin
APPLICANT: Kirsch, Donald R.
TITLE OF INVENTION: A DNA Sequence Encoding Sterol Delta
Patent No. 5512472
TITLE OF INVENTION: Reductase
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07470
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,131A
```

```
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/107,347
FILING DATE: 16-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gordon, Alan M.
REGISTRATION NUMBER: 30,637
REFERENCE/DOCKET NUMBER: 854-012 (32,141)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-831-3244
TELEFAX: 201-831-3305
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 419 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PUBLICATION INFORMATION:
AUTHORS: Worman, H. J.
AUTHORS: Evans, C. D.
AUTHORS: Blobel, G.
TITLE: The Lamin B Receptor of the Nuclear Envelope
JOURNAL: J. Cell Biol.
VOLUME: 111
PAGES: 1535-1542
DATE: 1990
RELEVANT RESIDUES IN SEQ ID NO: 3: FROM 190 TO 608
US-08-439-131A-3

Query Match      4.8%; Score 81; DB 1; Length 419;
Best Local Similarity 21.9%; Pred. No. 1.9;
Matches 39; Conservative 22; Mismatches 53; Indels 64; Gaps 7;

QY 58 IFFLGTALAFVLMPLHGRNLPFRSLSSWPFMLTALAVILONMAAHVFLETHD 117
DB 22 LMFLPATVL-YLVLMCKODDPSLWPFPLALESLWETKVGVELL-----WFF----- 70
QY 118 GHPOLTRNRVLYAATFLPLPLVLYGAMXXCSPSIARHP-TGY----- 162
DB 71 -----FQALFYLPPLPGKV-----EGLPLSNRKPQYINGFYAFILTA 111
QY 163 -----YTRNPLKIEVSQSHPAMTAFCSILLQASILPRTMAAPODSLRPG 208
DB 112 IOTLLYQFELHYLYDFVQPAVSAAAFSMALSTLYIT-----RSUKAPEEDLAPG 162
```

Search completed: March 13, 2003, 11:31:39
Job time: 22.7462 secs

PR 01-SEP-2000; 2000US-229780P.
PR 01-SEP-2000; 2000US-229848P.
PR 01-SEP-2000; 2000US-229850P.
PR 01-SEP-2000; 2000US-229850P.
PR 22-JAN-2001; 2001US-263337P.
PR 31-JAN-2001; 2001US-265518P.
PR 15-MAR-2001; 2001US-276451P.
PR 27-MAR-2001; 2001US-279196P.
PR 24-AUG-2001; 2001US-0393398.
XX
XX
XX (CURA-) CURAGEN CORP.
XX
PI Gerlach V, MacDougall JR, Smithson G, Stone DJ, Ellerman K;
PI Spytek KA, Zernussen BD, Rastelli L, Verney CAM, Patturajan M;
PI Tchernev VT, Padigaru M, Taupier RJ;
XX
DR WPI: 2002-293064/33.
DR N-PSDB; ABR55561.
XX
XX New isolated cytoplastic, nuclear, membrane bound and secreted
PT polypeptides, termed NOV, useful for treating inflammation, autoimmune
PT disorders, haemophilia, Lesch-Nyhan Syndrome, pancreatitis,
PT musculoskeletal disorders -
XX
XX Claim 1; Page 18; 245pp; English.

The invention relates to an isolated cytoplasmic nuclear membrane bound or secreted polypeptide, designated NOVX (actually NOV1, 2a, 2b, 3a, 3b, 4, 5a, 5b, 5c, 5d, 5e, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOVX, a mature form, or a variant of the mature form of NOVX. Also included are a polynucleotide encoding NOVX (or its complement), a vector comprising the polynucleotide, a cell comprising the vector, an anti-NOVX antibody, determining the presence of NOVX in a sample using the antibody, determining the presence of NOVX polynucleotide in a sample using a probe which binds to NOVX polynucleotide, identifying a agent which binds to NOVX (including modulators of NOVX). NOVX, the polynucleotide and the antibody are useful for diagnosing, treating or preventing a NOVX-associated disorder selected from cardiomyopathy, atherosclerosis, diabetes, a disorder related to cell signal processing and metabolic pathway modulation, inflammation, autoimmune disorders, scleroderma, transplantation, allergies, systemic lupus erythematosus, hemophilia, graft versus host disease, Alzheimer's disease, stroke, Leisch-Nyhan syndrome, peridontitis, pancreatitis, musculoskeletal disorders, Parkinson's disease, Huntington's disease, behavioural disorders, pain, neurodegenerative and neuropsychiatric disorders, hypertension, wound healing, obesity, growth and reproductive disorders, lung diseases and many other diseases and disorders listed in the specification. NOVX, the polynucleotide and the antibody are useful in screening assays, detection assays (e.g., chromosomal mapping, tissue typing, forensic biology), predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials and pharmacogenomic), and in methods of treatment (e.g., therapeutic and prophylactic). NOVX is useful as immunogen to produce antibodies immunospecific for NOVX, as vaccines to screen for potential agonist and antagonist compounds, and as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide is useful in gene therapy, to express NOVX, to detect NOVX mRNA or a genetic lesion in a NOVX gene, and to modulate NOVX activity. The vector is useful for producing non-human transgenic animals. The antibody is useful for isolating, and purifying NOVX and to monitor protein levels in tissue as part of a clinical testing procedure. The present sequence represents a NOVX protein.

Sequence 660 AA:

Query Match	66.9%	Score 1134	DB 23	Length 660
Best Local Similarity	76.9%	Pred. No. 5e-109		
Matches 237	Conservative 4	Mismatches 25	Indels 42	Gaps 5

```

QY      1  CYSALVASCCLTFVLMRSVLTNR-----LGSGSSD-----GQSFNMLPS 42
Db      353  CYSALVASCCLTFVLMRSVLTNRMLRALHRCMAIDLSPHRSPPHSPROAIFCMMSFS 411
QY      43  VPLP-----LPLAGLLVQOIIFFLGTTALAFVLMVVLHGRLLLPFSLSSNPFWLLTL 97

```

Db	413	AYQTAFCLEPRLPAGLVVQOIIFLGGTTALAFVLVMPVLVHGRNLLFRSLDESSMPFLTL	4727
QY	98	ALAVYLQNMMAHWVPLETHDGHPOLTNRRVLYAATFLLPLPLNVLGAMXXXC-----	1494
Db	473	ALAVYLQNMMAHWVPLETHDGHPOLTNRRVLYAATFLLPLPLNVLGAMVATWRVLSALY	5322
QY	150	-----SPSIAIRHP-----TPGYTYTRNPLKIEVSQSHPMATFCSLLLOAOSLLPRTM	1389
Db	533	NAIHIGQMDLSLEPRRAATLDPGYTYTRNPLKIEVSQSHPMATFCSLLLOAOSLLPRTM	5522
QY	199	AAPODSLRPGSEDEGKMLQTKDSMAGARGAARGARGLAVYLLHNPTLOVFRKTAL	2589
Db	593	AAPODSLRPGSEDEGKMLQTKDSMAGARGAARGARGLAVYLLHNPTLOVFRKTAL	6522
QY	259	LGANGAQP 266	
Db	653	LGANGAQP 660	

CC	RESULT 2
CC	AAB94108
ID	AAB94108 standard; Protein; 560 AA.
XX	
AC	AAB94108;
XX	
DT	26-JUN-2001 (first entry)
DE	Human protein sequence SEQ ID NO:14340.
XX	
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX	
OS	Homo sapiens.
PN	EPI074617-A2.
PD	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-0116126.
XX	
PR	29-JUL-1999; 99JP-0248036.
PR	27-AUG-1999; 99JP-0300253.
PR	11-JAN-2000; 2000JP-0118776.
PR	02-MAY-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
XX	
PA	(HELI-) HELIX RES INST.
PI	
PI	Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI	Izumi S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR	WPI; 2001-318749/34.
XX	
PT	Primer sets for synthesizing polynucleotides, particularly the 5602
PT	full-length cDNAs defined in the specification, and for the detection
PT	and/or diagnosis of the abnormality of the proteins encoded by the
PT	full-length cDNAs -
PS	
XX	Claim 8; SEQ ID 14340; 2537bp + CD ROM; English.
XX	
CC	The present invention describes primer sets for synthesizing 5602
CC	full-length cDNAs defined in the specification. Where a primer set
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC	to the complementary strand of a polynucleotide which comprises one of
CC	the 5602 nucleotide sequences defined in the specification, where the
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC	of an oligonucleotide comprising a sequence complementary to the
CC	complementary strand of a polynucleotide which comprises a 5'-end
CC	sequence and an oligonucleotide comprising a sequence complementary to a
CC	polynucleotide which comprises a 3'-end sequence, where the
CC	oligonucleotide comprises at least 15 nucleotides, and the combination of
CC	the 5'-end sequence/3'-end sequence is selected from those defined in
CC	the specification. The primer sets can be used in antisense therapy and
CC	in gene therapy. The primers are useful for synthesizing polynucleotides

CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 560 AA;

Query Match 65.5%; Score 1110.5; DB 22; Length 560;

Best Local Similarity 76.6%; Pred. No. 1.1e-106;

Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYIALVLSCLTFVLMRSLSVTR-----LGSGGSGD-----GQFSWNLFS 42
 DB 258 CYIALVLSCLTFVLMRSLSVTRTNLRALHRAALDLSLHSPHRSQAIFCWSFS 317
 QY 43 VPLPLPLAGLLVVOOIIFFLGTALAPLVMPVLHGRNLLFFRSLESSWPMWTLALAVI 102
 DB 318 AVQNAFTCLGLVVOOIIFFLGTALAPLVMPVLHGRNLLFFRSLESSWPMWTLALAVI 377
 QY 103 LQNNAAHWVLETHDGHQLTNRRLVYATFLPPLNVLVGAKXXC----- 149
 DB 378 LQNNAAHWVLETHDGHQLTNRRLVYATFLPPLNVLVGAWATWVLLSALYNAIL 437
 QY 150 -SPSLAIRHP-----TPGYTYRNPLKTEVSQSHAMTAFCSLLLOAQSLPRTMAAPD 203
 DB 438 GQMDLSLPRAATLDPEGYTYRNPLKTEVSQSHAMTAFCSLLLOAQSLPRTMAAPD 497
 QY 204 SLRGEDEGNQLQTKDSMAKGPAGKRGARWGLAYTLLHNPLOVFKRTALLGANG 263
 DB 498 SLRGEDEGNQLQTKDSMAKGPAGKRGARWGLAYTLLHNPLOVFKRTALLGANG 557
 QY 264 AQP 266
 DB 558 AQP 560

RESULT 3
 AAU04557
 ID AAU04557 standard; Protein; 667 AA.

XX AC AAU04557;

XX DT 26-SEP-2001 (first entry)

XX DE Human Stra6 homologue, PRO10282.

XX KW Human; Stra6; PRO10282; 15q23; tumour; breast cancer;
 KW head tumour; leukaemia; lymphoid malignancy; inflammatory disorder;
 KW immunogenic disorder; antigen; antibody.

XX OS Homo sapiens.

XX FH Key location/Qualifiers

FT Modified-site 8..12 /note= "Asn is N-glycosylated"

FT Region 50..56 /label= N_myristoylation_site

FT Domain 54..69 /label= Transmembrane_domain

FT Region 89..97 /note= "Region deleted in Stra6 variant PRO19578"

FT Domain 102..119 /label= Transmembrane_domain

FT Region 132..140 /note= "ATP/GTP-binding site motif A (P-loop)"

FT Domain 148..166 /label= Transmembrane_domain

FT Region 176..182 /label= N_myristoylation_site

FT FT

FT Domain 207..222 /label= Transmembrane_domain
 FT Region 241..247 /label= N_myristoylation_site
 FT Domain 301..320 /label= Transmembrane_domain
 FT Region 317..323 /label= N_myristoylation_site
 FT Region 341..347 /label= N_myristoylation_site
 FT Region 364..375 /note= "Prokaryotic membrane lipoprotein attachment site"
 FT 364..380 /label= Transmembrane_domain
 FT 431..451 /label= Transmembrane_domain
 FT 474..489 /label= Transmembrane_domain
 FT 525..531 /label= N_myristoylation_site
 FT 560..535 /label= Transmembrane_domain
 FT 627..633 /label= N_myristoylation_site
 FT 631..637 /label= N_myristoylation_site
 FT 640..646 /label= N_myristoylation_site
 FT 661..667 /label= N_myristoylation_site
 FT Region
 XX PN W0200151635-A2.
 XX PD 19-JUL-2001.
 XX PF 11-JAN-2001; 2001MO-US00847.
 XX PR 13-JAN-2000; 2000US-0175849.
 XX PR 14-APR-2000; 2000US-0197089.
 XX PR 29-AUG-2000; 2000US-0228914.
 XX PA (GETH) GENENTECH INC.
 XX PI Pennica D, Smith V, Wood WJ;
 XX DR WPI: 2001-442146/47.
 XX DR N-PSDB; AAS08630.
 XX PT Human nucleic acid encoding a PRO10282 or PRO19578 polypeptide (the PRO
 PT polypeptides are referred to as Stra6 polypeptides), useful in
 PT molecular biology, including use as hybridisation probes, and in
 PT chromosome and gene mapping -
 XX PS Claim 24; Fig 2; 159pp; English.
 XX CC The sequence is human PRO10282, a homologue of murine Stra6, a
 CC retinoic acid responsive protein thought to play an important role in
 CC early dorsoventral limb patterning during development and later in the
 CC control of endochondral ossification. The gene for the Stra6 homologue
 CC is located on chromosome 15q23. The Stra6 polypeptides, agonists,
 CC antagonists or anti-Stra6 antibodies are useful for preparation of a
 CC medicament useful in the treatment of a condition which is responsive to
 CC the Stra6 polypeptides, agonists, antagonists or anti-Stra6 antibodies.
 CC The Stra6 polypeptides may also be employed as molecular weight markers
 CC for protein electrophoresis. The Stra6 nucleic acids have applications in
 CC molecular biology, including use as hybridisation probes, and in
 CC chromosome and gene mapping. The antibodies and other anti-tumour
 CC compounds may be used to treat various conditions, including those
 CC characterised by overexpression and/or activation of amplified genes.
 CC Exemplary conditions or disorders include benign or malignant tumours
 CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
 CC colorectal, prostate, pancreatic, lung, vulval, thyroid, hepatic


```

KW pericyte cell proliferation; chondrocyte cell proliferation;
XX tumour necrosis factor-alpha.
XX Homo sapiens.
XX
XX WO200208288-A2.
XX
XX 31-JAN-2002.
XX
XX 29-JUN-2001; 2001WO-US21066.
XX
XX 20-JUL-2000; 2000US-219556P.
XX 25-JUL-2000; 2000US-220585P.
XX 25-JUL-2000; 2000US-220605P.
XX 25-JUL-2000; 2000US-220607P.
XX 25-JUL-2000; 2000US-220624P.
XX 25-JUL-2000; 2000US-220638P.
XX 25-JUL-2000; 2000US-220664P.
XX 25-JUL-2000; 2000US-220666P.
XX 26-JUL-2000; 2000US-220893P.
XX 28-JUL-2000; 2000WO-US20710.
XX 23-AUG-2000; 2000WO-US23522.
XX 24-AUG-2000; 2000WO-US23328.
XX 15-SEP-2000; 2000US-000000P.
XX 10-NOV-2000; 2000WO-US30873.
XX 28-NOV-2000; 2000US-253646P.
XX 01-DEC-2000; 2000WO-US32678.
XX 20-DEC-2000; 2000US-0747259.
XX 20-DEC-2000; 2000WO-US34956.
XX 28-FEB-2001; 2001WO-US06520.
XX 10-MAY-2001; 2001US-0854280.
XX 25-MAY-2001; 2001WO-US17092.
XX
XX (GENTH ) GENENTECH INC.
XX
XX Baker KP, Deenoyers L, Gerritsen ME, Goddard A, Godowski PJ;
XX Grimaldi JC, Gurney AL, Smith V, Stephan JF, Watanabe CK, Wood WI;
XX WPI; 2002-172001/22.
XX
XX N-PSDB; ABK33575.
XX
XX One hundred and twenty two nucleic acids encoding PRO polypeptides,
XX useful for treating a PRO related disorder and for diagnosing tumours
XX such as lung cancer, colon cancer, breast tumour, prostate tumour, rectal
XX tumour or liver tumour .
XX
XX Claim 11; Figure 80; 359bp; English.
XX
XX The invention relates to one hundred and twenty two nucleic acids
XX encoding PRO polypeptides. The sequences of the 122 PRO polynucleotides
XX encode human secreted proteins. The PRO nucleic acids, polypeptides,
XX agonists and antagonists are useful for treating a PRO related disorder.
XX The PRO polypeptides are useful for diagnosing tumours, especially lung
XX cancer, colon cancer, breast tumour, prostate tumours, rectal tumour or
XX liver tumour. The PRO polypeptides are useful for stimulating the
XX proliferation of, or gene expression in, pericyte cells, for stimulating
XX the proliferation or differentiation of chondrocyte cells, for
XX stimulating the release of tumour necrosis factor-alpha from human blood,
XX for stimulating or inhibiting the proliferation of normal human dermal
XX fibroblast cells. The PRO polypeptide may also be used as molecular
XX weight markers and for tissue typing. The PRO nucleic acids have
XX applications in molecular biology, including use as hybridisation probes,
XX and in chromosome and gene mapping. AU83592-AU83713 represent human PRO
XX protein sequences of the invention.
XX
XX Sequence 667 AA:
XX
XX Query Match 65.5%; Score 1110.5; DB 23; Length 667;
XX Best Local Similarity 76.6%; Pred. No. 1.4e-106;
XX Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4
XX
XX 1 CTTAALVLSCLTFLVLMNSLYTHR-----LGGGSGD-----GQFSWMLPS 42
XX |||||

```

Db	365	CYLSALVLSCLLTPLVLMRSLVYTRTLRALHRCALDLSPLHRSHPERQAI	PCMSFS	424
Qy	43	VPLPLPLAGLLVQOIIFFLGTTALAFVLVMPVLHGNNLFFRSLESSMPFWTLALAVI		102
Db	425	AYQTAFCICGLLVQOIIFFLGTTALAFVLVMPVLHGNNLFFRSLESSMPFWTLALAVI		484
Qy	103	LQNMAAHVPLETHDGHQQLNRRVLYAATFLLPFLNVLYGAKXXXC-----		149
Db	485	LQNMAAHVPLETHDGHQQLNRRVLYAATFLLPFLNVLYGAKXXXC-----		149
Qy	150	-SPSIAIRHP-----TRGYTYRNFLEKIEVSQSHPMATAFCSLLLOAQSLPRITMAPOD		203
Db	545	QGMDISLLPRPAATLIDGYTYRNFLEKIEVSQSHPMATAFCSLLLOAQSLPRITMAPOD		604
Qy	204	SLRPEEBEGMQLLOTQDSMAKGRAPPAKGRARMGATYLLNHPLOYFRKATLIGANG		263
Db	605	SLRPEEBEGMQLLOTQDSMAKGRAPPAKGRARMGATYLLNHPLOYFRKATLIGANG		664
Qy	264	AQP 266		
Db	665	AQP 667		
RESULT 7				
ABB05048	ID	ABB05048	standard; Protein; 667 AA.	
XX	AC	ABB05048;		
XX	DT	25-MAR-2002	(first entry)	
XX	DE	Human NOV8c protein SEQ ID NO:32.		
KM	Human; NOVX; cytostatic; antidiabetic; anorectic; antibacterial;			
KM	fungicide; virucide; procoagulant; analgesic; antiparkinsonian;			
KM	anticholinergic; hypotensive; osteopathic; antiinflammatory; antiulcer;			
KM	neuroprotective; cardiac; antiallergic; antidepressant; nootropic;			
KM	anticonvulsant; neuroleptic; gene therapy; vaccine; immune disease;			
KM	developmental disease; taste and scent detectability disorder; infection;			
KM	Burkitt's lymphoma; signal transduction pathway disorder; pain; cancer;			
KM	retinal disease; feeding disorder; asthma; Parkinson's disease; ulcer;			
KM	noninsulin-dependent diabetes mellitus; acute heart failure; hypotension;			
KM	hyperextension; urinary retention; osteoporosis; Crohn's disease; allergy;			
KM	benign sclerotic; angina pectoris; myocardial infarction; delirium;			
KM	multiple sclerosis; hypertrophy; manic depression; dementia; dyskinesia;			
KM	severe mental retardation; Huntington's disease; chromosome 15;			
KM	Gilles de la Tourette syndrome.			
XX				
OS	Homo sapiens.			
XX				
XX	MO200190155-A2.			
XX				
PD	29-NOV-2001.			
XX				
XX	24-MAY-2001; 2001WO-US17073.			
XX				
XX	24-MAY-2000; 2000US-206679P.			
PR	24-MAY-2000; 2000US-206688P.			
PR	24-MAY-2000; 2000US-206829P.			
PR	30-MAY-2000; 2000US-207748P.			
PR	30-MAY-2000; 2000US-207798P.			
PR	31-MAY-2000; 2000US-208263P.			
PR	02-JUN-2000; 2000US-208831P.			
PR	05-JUN-2000; 2000US-209451P.			
PR	07-JUN-2000; 2000US-210060P.			
PR	20-JUL-2000; 2000US-219507P.			
PR	26-JUL-2000; 2000US-221337P.			
PR	31-JUL-2000; 2000US-221927P.			
PR	19-JAN-2001; 2001US-263135P.			
PR	24-JAN-2001; 2001US-263688P.			
PR	24-JAN-2001; 2001US-263694P.			
PR	23-MAY-2001; 2001US-086376P.			
XX				

PA (CURA-) CURAGEN CORP.
 XX
 PI Spletak KA, Majumder K, Tchernev VT, Mishra V, Padigar M;
 XX Spaderma SK, Shenoy S, Rastelli L, Li L, Taupier RJ, Gangoli E;
 XX WPI; 2002-106174/14.
 DR N-PSDB; ABA92669.
 XX
 XX Novel polypeptide, useful for treating pain, cancer, urinary retention,
 PT osteoporosis, Crohn's disease, diabetes, acute heart failure, dementia,
 PT asthma, ulcer, allergy and Huntington's disease, comprises isolated
 PT polypeptide NOVX -
 PS
 XX Claim 1; Page 106; 266pp; English.
 CC The present invention describes human NOVX proteins (NOV1 to NOV9). NOVX
 CC proteins (I) have cytosolic, antidiabetic, anorectic, antibacterial,
 CC fungicide, virucide, protozoacide, analgesic, antiparkinsonian,
 CC antileptemic, hypotensive, osteopathic, antiinflammatory, antiulcer,
 CC neuroprotective, cardiant, antiallergic, antidepressant, nootropic,
 CC anticonvulsant and neuroleptic activities. (I) and polynucleotides (II)
 CC can be used in gene therapy and vaccine production. (I) and (II) can be
 CC used for treating or preventing a NOVX-associated disorder such as
 CC cardiomyopathy, atherosclerosis and diabetes in a human, where the
 CC disorder is related to cell signal processing and metabolic path way
 CC modulation, in a subject, preferably human. (I) and (II) can be used for
 CC diagnosing, preventing or treating developmental diseases, immune
 CC diseases, taste and scent detectability disorder, Burkitt's lymphoma,
 CC signal transduction pathway disorders, retinal diseases including those
 CC involving photoreception, cell growth rate disorders, feeding disorders,
 CC noninsulin-dependent diabetes mellitus, infections, pain, cancer, asthma,
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,
 CC urinary retention, osteoporosis, Crohn's disease, multiple sclerosis,
 CC Albright Hereditary Osteodystrophy, angina pectoris, ulcers, myocardial
 CC infarction, allergies, benign prostatic hypertrophy, manic depression,
 CC delirium, dementia, severe mental retardation and dyskinesia, such as
 CC Huntington's disease or Gilles de la Tourette syndrome. The present
 CC sequence represents the human NOV8c protein from the present invention.
 CC Human NOV2a is isolated to chromosome 15.
 XX
 XX Sequence 667 AA;
 SQ
 Query Match 65.5%; Score 1110.5; DB 23; Length 667;
 Best Local Similarity 76.6%; Pred. No. 1.4e-106;
 Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;
 QY 1 CYTALVLSCLTFLVLMRSIVTR-----LGGSGSGD-----GQSSMNF 42
 DB 365 CYTALVLSCLTFLVLMRSIVTR-----LGGSGSGD-----GQSSMNF 424
 QY 43 VPLPLPLAGLVVQOIIFFLGTALLAFVLMPLVLRLLPFRSLDESSPMLTALAVI 102
 DB 425 AYQAFICLGLVQOIIFFLGTALLAFVLMPLVLRLLPFRSLDESSPMLTALAVI 484
 QY 103 LQNNAAHVLFLETHDGHQPLTNRRLVYATFLPPLNVLVGMXXXC----- 149
 DB 485 LQNNAAHVLFLETHDGHQPLTNRRLVYATFLPPLNVLVGMVATWRLVLSALYNAIL 544
 QY 150 -SPSIALRHP-----TTCYTYRRNFKLEVGOSHPAMAFSGLLIOASLLPRTMAAOD 203
 DB 545 GOMLPLPRAATLDPCYTYRRNFKLEVGOSHPAMAFSGLLIOASLLPRTMAAOD 604
 QY 204 SLRGEDEGQQLQTKOSMAKRGARPGAXRGRARGLAYTLLHNPTLVPRKTALGANG 263
 DB 605 SLRGEDEGQQLQTKOSMAKRGARPGAXRGRARGLAYTLLHNPTLVPRKTALGANG 664
 QY 264 AQP 266
 DB 665 AQP 667

ID AAU91278 standard; Protein; 689 AA.
 XX
 AC AAU91278;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Human NOV2b protein.
 XX
 KM Human; NOVX; gene therapy; cardiomyopathy; atherosclerosis;
 KM diabetes; cell signal processing; metabolic pathway modulation;
 KM inflammation; autoimmune disorder; scleroderma; transplantation;
 KM allergy; systemic lupus erythematosus; haemophilia; Alzheimer's disease;
 KM graft versus host disease; Leech-Nyhan syndrome; periodontitis;
 KM pancreaticitis; musculoskeletal disorder; Parkinson's disease;
 KM Huntington's disease; behavioural disorder; pain; obesity; wound healing;
 KM neurodegenerative disorder; neuropsychiatric disorder; hypertension;
 KM growth disorder; reproductive disorder; lung disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200216600-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 27-AUG-2001; 2001WO-US26518.
 XX
 PR 25-AUG-2000; 2000US-227800P.
 PR 25-AUG-2000; 2000US-228205P.
 PR 25-AUG-2000; 2000US-228324P.
 PR 30-AUG-2000; 2000US-228977P.
 PR 30-AUG-2000; 2000US-229180P.
 PR 01-SEP-2000; 2000US-229780P.
 PR 01-SEP-2000; 2000US-229848P.
 PR 01-SEP-2000; 2000US-229850P.
 PR 22-JAN-2001; 2001US-263337P.
 PR 31-JAN-2001; 2001US-265318P.
 PR 15-MAR-2001; 2001US-276451P.
 PR 27-MAR-2001; 2001US-279196P.
 PR 24-AUG-2001; 2001US-0393398.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 PI Gerlach V, MacDougall JR, Smithson G, Stone DJ, Ellerman K;
 PI Spletak KA, Zernusen BD, Rastelli L, Verney CAM, Patrujan M;
 PI Tchernev VT, Padigar M, Taupier RJ;
 XX
 DR WPI; 2002-292064/33.
 DR N-PSDB; ABK55562.
 XX
 PT New isolated cytoplasmic, nuclear, membrane bound and secreted
 PT polypeptides, termed NOVX, useful for treating inflammation, autoimmune
 PT disorders, haemophilia, Leech-Nyhan syndrome, pancreatitis,
 PT musculoskeletal disorders -
 XX
 XX Claim 1; Page 20; 245pp; English.
 XX
 PS The invention relates to an isolated cytoplasmic, nuclear, membrane bound
 CC or secreted polypeptide, designated NOVX (actually NOV1, 2a, 2b, 3a, 3b,
 CC 4, 5a, 5b, 5c, 5d, 5e, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOVX, a
 CC mature form, or a variant of the mature form of NOVX. Also included
 CC are a polynucleotide encoding NOVX (or its complement), a vector
 CC comprising the polynucleotide, a cell comprising the vector, an
 CC anti-NOVX antibody, determining the presence of NOVX in a sample
 CC using the antibody, determining the presence of NOVX polynucleotide in
 CC a sample using a probe which binds to NOVX polynucleotide, identifying a
 CC an agent which binds to NOVX (including modulators of NOVX). NOVX, the
 CC polynucleotide and the antibody are useful for diagnosing, treating
 CC or preventing a NOVX-associated disorder selected from cardiomyopathy,
 CC atherosclerosis, diabetes, a disorder related to cell signal processing,
 CC and metabolic pathway modulation, inflammation, autoimmune disorders,
 CC scleroderma, transplantation, allergies, systemic lupus erythematosus,
 CC haemophilia, graft versus host disease, Alzheimer's disease, stroke,
 CC Leech-Nyhan syndrome, periodontitis, pancreatitis, musculoskeletal

CC disorders, Parkinson's disease, Huntington's disease, behavioural
 CC disorders, pain, neurodegenerative and neuropsychiatric disorders,
 CC hyperaesthesia, wound healing, obesity, growth and reproductive
 CC disorders, lung diseases and many other diseases and disorders listed in
 CC the specification. NOVX, the polynucleotide and the antibody are useful
 CC in screening assays, detection assays (e.g., chromosomal mapping, tissue
 CC typing, forensic biology), predictive medicine (e.g., diagnostic assays,
 CC prognostic assays, monitoring clinical trials and pharmacogenomic), and
 CC in methods of treatment (e.g., therapeutic and prophylactic). NOVX is
 CC useful as immunogen to produce antibodies immunospecific for NOVX, as
 CC vaccines to screen for potential agonist and antagonist compounds, and
 CC as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide
 CC is useful in gene therapy, to express NOVX, to detect NOVX mRNA
 CC or a genetic lesion in a NOVX gene, and to modulate NOVX activity. The
 CC vector is useful for producing non-human transgenic animals. The antibody
 CC is useful for isolating, and purifying NOVX and to monitor protein levels
 CC in tissue as part of a clinical testing procedure. The present sequence
 CC represents a NOVX protein.

XX Sequence 689 AA;

Query Match 65.5%; Score 1110.5; DB 23; Length 689;
 Best Local Similarity 76.6%; Pred. No. 1.5e-106;
 Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYISALVLSCLTFLVLRSLVTHR-----LGSQSGSD-----GQFSMNFSS 42
 DB 387 CYISALVLSCLTFLVLRSLVTHRSLRALHGAALDLSPLHRSRPHSRQAIFCMMSFS 446
 QY 43 VPLPLPLAGLVVOOIFPLGTTALAFVLPVLMVHGRNLLPFRSLESMPWLTALAVI 102
 DB 447 AVQTAFLICLGLVOOIFPLGTTALAFVLPVLMVHGRNLLPFRSLESMPWLTALAVI 506
 QY 103 LQNMAMHVPLETHDGHQPLNRRVLYAATFLPLPLVNVGAMXXC----- 149
 DB 507 LQNMAMHVPLETHDGHQPLNRRVLYAATFLPLPLVNVGAMVATWVLLSALYNATHL 566
 QY 150 -SPSIAIRHP-----TPGYTYRNFLKIEVSQSHPMATFCSLLQAOSLLPTMAAPD 203
 DB 567 GQMDLSLPLPAAATLDPEYTYRNFLKIEVSQSHPMATFCSLLQAOSLLPTMAAPD 626
 QY 204 SLRGEEDBEGHQLQTKSMAGKAPGAXRGRARGLAYTLINPTLOVFRKTALLGANG 263
 DB 627 SLRGEEDBEGHQLQTKSMAGKAPGASRGRARGLAYTLINPTLOVFRKTALLGANG 686
 QY 264 AQP 266
 DB 687 AQP 689

RESULT 9
 ID AAU04558 standard; Protein: 658 AA.

XX AAU04558;
 XX 26-SEP-2001 (first entry)
 XX Human Stra6 homologue, PRO19578.
 DE Human Stra6 homologue, PRO19578.
 XX Human; Stra6, PRO19578, 15q23; tumour; breast cancer;
 KM head tumour; leukaemia; lymphoid malignancy; inflammatory disorder;
 KM immunogenic disorder; antigen; antibody.
 XX

OS Homo sapiens.
 XX Key Location/Qualifiers
 FH Modified-site 8..12
 FT /note= "Aan is N-glycosylated"
 FT Region 50..56
 FT /label= N_myristoylation_site
 FT Domain 54..71
 FT /label= Transmembrane_domain

FT Domain 93..111
 FT /label= Transmembrane_domain
 FT Region 123..131
 FT /note= "ATP/GTP-binding site motif A (P-loop)"
 FT Domain 140..157
 FT /label= Transmembrane_domain
 FT Region 167..173
 FT /label= N_myristoylation_site
 FT Domain 197..214
 FT /label= Transmembrane_domain
 FT Region 232..238
 FT /label= N_myristoylation_site
 FT Domain 291..312
 FT /label= Transmembrane_domain
 FT Region 308..314
 FT /label= N_myristoylation_site
 FT Region 332..338
 FT /label= N_myristoylation_site
 FT Region 355..366
 FT /note= "Prokaryotic membrane lipoprotein attachment site"
 FT Domain 356..371
 FT /label= Transmembrane_domain
 FT Domain 425..444
 FT /label= Transmembrane_domain
 FT Domain 464..481
 FT /label= Transmembrane_domain
 FT Domain 505..522
 FT /label= Transmembrane_domain
 FT Region 516..522
 FT /label= N_myristoylation_site
 FT Region 618..624
 FT /label= N_myristoylation_site
 FT Region 622..628
 FT /label= N_myristoylation_site
 FT Region 631..637
 FT /label= N_myristoylation_site
 FT Region 652..658
 FT /label= N_myristoylation_site
 XX NO200151635-A2.
 XX 19-JUL-2001.
 XX 11-JAN-2001; 2001WO-US00847.
 XX 13-JAN-2000; 2000US-0175849.
 XX 14-APR-2000; 2000US-0197089.
 XX 29-AUG-2000; 2000US-0228914.
 XX (GENTH) GENENTECH INC.
 XX Pennica D, Smith V, Wood WI;
 XX WPI; 2001-442146/47.
 XX DR N-PSDB; AAS08632.
 XX Human nucleic acid encoding a PRO10282 or PRO19578 polypeptide (the PRO
 PT polypeptides are referred to as Stra6 polypeptides), useful in
 PT molecular biology, including use as hybridisation probes, and in
 PT chromosome and gene mapping -
 PS Claim 24; Fig 7; 15pp; English.
 XX The sequence is human PRO19578 (a variant possibly due to
 CC alternative splicing of the Stra6 mRNA), a homologue of murine Stra6, a
 CC retinoic acid responsive protein thought to play an important role in
 CC early dorsoventral limb patterning during development and later in the
 CC control of endochondral ossification. The gene for the Stra6 homologue
 CC is located on chromosome 15q23. The Stra6 polypeptides, agonists,
 CC antagonists or anti-Stra6 antibodies are useful for preparation of a
 CC medicament useful in the treatment of a condition which is responsive to
 CC the Stra6 polypeptides, agonists, antagonists or anti-Stra6 antibodies.

CC administered for the treatment of various disorders or conditions,
 CC including those characterised by overexpression and/or activation of the
 CC tumour antigens, where such conditions or disorders include benign or
 CC malignant tumours; leukaemias and lymphoid malignancies; neuronal, glial,
 CC astrocytal, hypochalamic and other glandular, macropagial, epithelial,
 CC stromal and blastocellic disorders. The present sequence represents
 CC the human Stra6 protein variant (PRO19578) which is actively upregulated
 CC by retinoid using the method of the invention. The gene encoding the
 CC human Stra6 protein is located on human chromosome 15q23.

XX Sequence 658 AA;

Query Match 65.3%; Score 1106.5; DB 23; Length 658;
 Best Local Similarity 76.2%; Pred. No. 3.6e-106;
 Matches 231; Conservative 5; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYISALVLSCLTFLVLMRSVLTNR-----LQSGGSGD-----GQFSWNLFS 42
 DB 356 CYISALVLSCLTFLVLMRSVLTNRRLRALHGRNALDLSPLHRSPPHSROAIFCWMGFS 415

QY 43 VPLPLPLAGLLVQOIIFFLGTALAPLVLMVPLHGRNLLFRSLSSGMPFLTALAVI 102
 DB 416 AYQTAFTCLGLVQOIIFFLGTALAPLVLMVPLHGRNLLFRSLSSGMPFLTALAVI 475

QY 103 LQNMMAHNVFLETHDGHQDLTNRRVLYAATFLPLVNLVGMXXXXC----- 149
 DB 476 LQNMMAHNVFLETHDGHQDLTNRRVLYAATFLPLVNLVGMIVATWVLSALYNAILH 535

QY 150 -SPSIAIHP-----TPGYTYRNLTKIEVSQSPHMTAFCSLLIQAGSLPRTMAAPD 203
 DB 536 GQMDLSLPPRAATLDPGYTYRNLTKIEVSQSPHMTAFCSLLIQAGSLPRTMAAPD 595

QY 204 SLRPEDEGQMLQTKDSMAKGARPGAXRGARAGLAYTLHNPFLQVFRKTALGANG 263
 DB 596 SLRPEDEGQMLQTKDSMAKGARPGAXRGARAGLAYTLHNPFLQVFRKTALGANG 655

QY 264 AQP 266
 DB 656 AQP 658

RESULT 11
 ABB05046
 ID ABB05046 standard; Protein; 653 AA.

AC ABB05046;
 XX 25-MAR-2002 (first entry)

DE Human NOV8a protein SEQ ID NO:28.

XX Human; NOVX; cytostatic; antidiabetic; anorectic; antibacterial;
 KM fungicide; virucide; protozoacide; analgesic; antiparkinsonian;
 KM antiaesthetic; hypotensive; osteopathic; antiinflammatory; antitumor;
 KM antineoplastic; cardiatic; antiallergic; antidepressant; nootropic;
 KM anticonvulsant; neuroleptic; gene therapy; vaccine; immune disease;
 KM developmental disease; taste and scent detectability disorder; infection;
 KM Burkitt's lymphoma; signal transduction pathway disorder; pain; cancer;
 KM retinal disease; feeding disorder; asthma; Parkinson's disease; ulcer;
 KM noninsulin-dependent diabetes mellitus; acute heart failure; hypotension;
 KM hyperkalemia; urinary retention; osteoporosis; Crohn's disease; allergy;
 KM multiple sclerosis; angina pectoris; myocardial infarction; delirium;
 KM benign prostatic hypertrophy; manic depression; dementia; dyskinesia;
 KM severe mental retardation; Huntington's disease; chromosome 15;
 KM Gilles de la Tourette syndrome.

XX Homo sapiens.

XX WO200190155-A2.

XX 29-NOV-2001.

XX 24-MAY-2001; 2001WO-US17073.

XX 24-MAY-2000; 2000US-206679P.
 PR 24-MAY-2000; 2000US-206689P.
 PR 24-MAY-2000; 2000US-206829P.
 PR 30-MAY-2000; 2000US-207748P.
 PR 30-MAY-2000; 2000US-207798P.
 PR 31-MAY-2000; 2000US-208263P.
 PR 02-JUN-2000; 2000US-208831P.
 PR 05-JUN-2000; 2000US-209451P.
 PR 07-JUN-2000; 2000US-210060P.
 PR 20-JUL-2000; 2000US-219507P.
 PR 26-JUL-2000; 2000US-221337P.
 PR 31-JUL-2000; 2000US-221927P.
 PR 19-JAN-2001; 2001US-263135P.
 PR 24-JAN-2001; 2001US-263688P.
 PR 24-JAN-2001; 2001US-263694P.
 PR 23-MAY-2001; 2001US-0863776.
 XX (CURA-) CURAGEN CORP.
 XX Spytex KA, Majumder K, Tchernev VT, Mishra V, Padigaru M;
 PI Spaderna SK, Shenoy S, Rastelli L, Li L, Taupier RJ, Gangoli E;
 XX WPI; 2002-106174/14.
 DR N-PSDB; ABA92663.

PT Novel polypeptide, useful for treating pain, cancer, urinary retention,
 PT osteoporosis, Crohn's disease, diabetes, acute heart failure, dementia,
 PT asthma, ulcer, allergy and Huntington's disease, comprises isolated
 PT polypeptide NOVX -
 PS
 XX Claim 1; Page 98; 266pp; English.

XX The present invention describes human NOVX proteins (NOV1 to NOV9). NOVX
 CC proteins (I) have cytostatic, antidiabetic, anorectic, antibacterial,
 CC fungicide, virucide, protozoacide, analgesic, antiparkinsonian,
 CC antiaesthetic, hypotensive, osteopathic, antiinflammatory, antitumor,
 CC antineoplastic, cardiatic, antiallergic, antidepressant, nootropic,
 CC anticonvulsant and neuroleptic activities. (I) and polynucleotides (II)
 CC can be used in gene therapy and vaccine production. (I) and (II) can be
 CC used for treating or preventing a NOVX-associated disorder such as
 CC cardiomyopathy, atherosclerosis and diabetes in a human, where the
 CC disorder is related to cell signal processing and metabolic path way
 CC modulation, in a subject, preferably human. (I) and (II) can be used for
 CC diagnosing, preventing or treating developmental diseases, immune
 CC diseases, taste and scent detectability disorder, Burkitt's lymphoma,
 CC signal transduction pathway disorders, retinal diseases including those
 CC involving photoreception, cell growth rate disorders, feeding disorders,
 CC noninsulin-dependent diabetes mellitus, infections, pain, cancer, asthma,
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,
 CC urinary retention, osteoporosis, Crohn's disease, multiple sclerosis,
 CC Albrecht Hereditary Osteodystrophy, angina pectoris, ulcers, myocardial
 CC infarction, allergies, benign prostatic hypertrophy, manic depression,
 CC delirium, dementia, severe mental retardation and dyskinesia, such as
 CC Huntington's disease or Gilles de la Tourette syndrome. The present
 CC sequence represents the human NOV8a protein from the present invention.
 CC Human NOV8a is isolated to chromosome 15.

XX Sequence 653 AA;

Query Match 62.6%; Score 1060.5; DB 23; Length 653;
 Best Local Similarity 74.6%; Pred. No. 2.2e-101;
 Matches 226; Conservative 4; Mismatches 30; Indels 43; Gaps 6;

QY 1 CYISALVLSCLTFLVLMRSVLTNR-----LQSGGSGD-----GQFSWNLFS 42
 DB 357 CYISALVLSCLTFLVLMRSVLTNRRLRALHGRNALDLSPLHRSPPHSROAIFCWMGFS 416

QY 43 VPLPLPLAGLLVQOIIFFLGTALAPLVLMVPLHGRNLLFRSLSSGMPFLTALAVI 102
 DB 417 AYQTAFTCLGLVQOIIFFLGTALAPLVLMVPLHGRNLLFRSLSSGMPFLTALAVI 470

QY 103 LQNMMAHNVFLETHDGHQDLTNRRVLYAATFLPLVNLVGMXXXXC----- 149

DB 471 LQNNAAWVFLETHDGHQPLTNRRVLAATFLPPLNVLGAMVATWRVLLSALYNIAHL 530
 QY 150 -SPSIAIRHP-----TRGYTYRNFLKIEVSQSHPAMTAFCSLLLOQSLPRTMAAPD 203
 DB 531 GQMDLSLPPRAATLDGTYTYRNFLKIEVSQSHPAMTAFCSLLLOQSLPRTMAAPD 590
 QY 204 SLRGEDEGMQLLOTQDSMAKGRPGAGRGARWGLAYTLTLLNPTLQVFRKTXLLGANG 263
 DB 591 SLRGEDEGMQLLOTQDSMAKGRPGAGRGARWGLAYTLTLLNPTLQVFRKTXLLGANG 650
 QY 264 AQP 266
 DB 651 AQP 653

RESULT 12
 ABB05047
 ID ABB05047 standard; Protein: 662 AA.
 AC ABB05047;
 XX
 DT 25-MAR-2002 (first entry)
 XX
 DE Human NOV8b protein SEQ ID NO:30.
 XX
 KW Human; NOVX; cytostatic; antidiabetic; anorectic; antibacterial;
 KW fungicide; virucide; protozoacide; analgesic; antiparkinsonian;
 KW antiasmatic; hypotensive; osteopathic; antiinflammatory; antiulcer;
 KW neuroprotective; cardiac; antiallergic; antidepressant; nootropic;
 KW anticonvulsant; neuroleptic; gene therapy; vaccine; immune disease;
 KW developmental disease; taste and scent detectability disorder; infection;
 KW Burkitt's lymphoma; signal transduction pathway disorder; pain; cancer;
 KW retinal disease; feeding disorder; asthma; Parkinson's disease; ulcer;
 KW noninsulin-dependent diabetes mellitus; acute heart failure; hypotension;
 KW hypertension; urinary retention; osteoporosis; Crohn's disease; allergy;
 KW multiple sclerosis; angina pectoris; myocardial infarction; delirium;
 KW benign prostatic hypertrophy; manic depression; dementia; dyskinesia;
 KW severe metal retardation; Huntington's disease;
 KW Gilles de la Tourette syndrome.
 XX
 OS Homo sapiens.
 XX
 EN WO200190155-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 24-MAY-2001; 2001MO-US17073.
 XX
 PR 24-MAY-2000; 2000US-206679P.
 PR 24-MAY-2000; 2000US-206688P.
 PR 24-MAY-2000; 2000US-206829P.
 PR 30-MAY-2000; 2000US-207748P.
 PR 30-MAY-2000; 2000US-207798P.
 PR 31-MAY-2000; 2000US-208263P.
 PR 02-JUN-2000; 2000US-208831P.
 PR 05-JUN-2000; 2000US-209451P.
 PR 07-JUN-2000; 2000US-210060P.
 PR 20-JUL-2000; 2000US-219507P.
 PR 26-JUL-2000; 2000US-221337P.
 PR 31-JUL-2000; 2000US-221927P.
 PR 19-JAN-2001; 2001US-263135P.
 PR 24-JAN-2001; 2001US-263688P.
 PR 24-JAN-2001; 2001US-263694P.
 PR 23-MAY-2001; 2001US-0863776.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Spytek KA, Majumder K, Tchernyev VT, Mishra V, Padigaru M;
 PI Spaderma SK, Shenoy S, Rastelli L, Li L, Taupier RJ, Gangoli B;
 XX
 DR WPI: 2002-106174/14.
 DR N-PSDB; ABA92666.

XX Novel polypeptide, useful for treating pain, cancer, urinary retention,
 PT osteoporosis, Crohn's disease, diabetes, acute heart failure, dementia,
 PT asthma, ulcer, allergy and Huntington's disease, comprises isolated
 PT polypeptide NOVX -
 XX
 PS Claim 1; Page 103; 266pp; English.
 XX
 CC The present invention describes human NOVX proteins (NOV1 to NOV9). NOVX
 CC proteins (I) have cytostatic, antidiabetic, anorectic, antibacterial,
 CC fungicide, virucide, protozoacide, analgesic, antiparkinsonian,
 CC antiasmatic, hypotensive, osteopathic, antiinflammatory, antiulcer,
 CC neuroprotective, cardiac, antiallergic, antidepressant, nootropic,
 CC anticonvulsant and neuroleptic activities. (I) and polynucleotides (II)
 CC can be used in gene therapy and vaccine production. (I) and (II) can be
 CC used for treating or preventing a NOVX-associated disorder such as
 CC cardiomyopathy, atherosclerosis and diabetes in a human, where the
 CC disorder is related to cell signal processing and metabolic path way
 CC modulation, in a subject, preferably human. (I) and (II) can be used for
 CC diagnosing, preventing or treating developmental diseases, immune
 CC diseases, taste and scent detectability disorder, Burkitt's lymphoma,
 CC signal transduction pathway disorders, retinal diseases including those
 CC involving photoreception, cell growth rate disorders, feeding disorders,
 CC noninsulin-dependent diabetes mellitus, infections, pain, cancer, asthma,
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,
 CC urinary retention, osteoporosis, Crohn's disease, multiple sclerosis,
 CC Albright Hereditary Osteodystrophy, angina pectoris, ulcers, myocardial
 CC infarction, allergies, benign prostatic hypertrophy, manic depression,
 CC delirium, dementia, severe metal retardation and dyskinesias, such as
 CC Huntington's disease or Gilles de la Tourette syndrome. The present
 CC sequence represents the human NOV8b protein from the present invention.
 XX
 SQ Sequence 662 AA;
 XX
 Query Match 62.6%; Score 1060.5; DB 23; Length 662;
 Best Local Similarity 74.6%; Pred. No. 2,2e-101;
 Matches 226; Conservative 4; Mismatches 30; Indels 43; Gaps 6;
 QY 1 CYISALVLSCLTFPLVMSRLVTHR-----LGSQSGSD-----GQPSMVLFS 42
 DB 366 CYISALVLSCLTFPLVMSRLVTHRTVRLALHGAALDLSPLHRSPPRSQALFCWMSFS 425
 QY 43 VPLPLPPLAGLLVQOIIFFIGTALAFVLPVLMVPHGNLLFFRSLESSWPFMTLALAVI 102
 DB 426 AYCTAFICLGLVQOIIFFIGTALAFVLPVLMVPHGNLLFFRSLESSWPFMTLALAVI 102
 QY 103 LQNNAAWVFLETHDGHQPLTNRRVLAATFLPPLNVLGAMVATWRVLLSALYNIAHL 539
 DB 480 LQNNAAWVFLETHDGHQPLTNRRVLAATFLPPLNVLGAMVATWRVLLSALYNIAHL 539
 QY 150 -SPSIAIRHP-----TRGYTYRNFLKIEVSQSHPAMTAFCSLLLOQSLPRTMAAPD 203
 DB 540 GQMDLSLPPRAATLDGTYTYRNFLKIEVSQSHPAMTAFCSLLLOQSLPRTMAAPD 599
 QY 204 SLRGEDEGMQLLOTQDSMAKGRPGAGRGARWGLAYTLTLLNPTLQVFRKTXLLGANG 263
 DB 600 SLRGEDEGMQLLOTQDSMAKGRPGAGRGARWGLAYTLTLLNPTLQVFRKTXLLGANG 659
 QY 264 AQP 266
 DB 660 AQP 662

RESULT 13
 AAM23581
 ID AAM23581 standard; Protein: 247 AA.
 XX
 AC AAM23581;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Murine EST encoded protein SEQ ID NO: 1106.
 XX

KM Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KM colorectal carcinoma.
 XX
 XX Homo sapiens.
 OS
 XX WO200122920-A2.
 PN
 XX
 XX 05-APR-2001.
 PD
 XX 28-SEP-2000; 2000WO-US26524.
 PF
 XX 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;
 PI
 XX WPI; 2001-235357/24.
 XX N-PSDB; AAH33488.
 DR
 XX
 XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 XX Claim 11; Page 6606-6607; 9803pp; English.
 PS
 XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 CC
 XX
 XX
 SQ Sequence 178 AA;

Query Match 28.6%; Score 485; DB 22; Length 178;
 Best Local Similarity 88.2%; Pred. No. 3.1e-42;
 Matches 97; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

QY 36 FSNWLFSEVPLPLPPLAGLVVQIIFFLGTALAFVLMPVVHGRNLLFFRSLSSMPWL 95
   |||||
DB 27 FCMNSFSAVOTAFICLGILVQIIFFLGTALAFVLMPVVHGRNLLFFRSLSSMPWL 86
   |||||
QY 96 TLALAVILQNNAAAHVPLLETHDGHPOLNRRVLYAATFLFPLNVVGAM 145
   |||||
DB 87 TLALAVILQNNAAAHVPLLETHDGHPOLNRRVLYAATFLFPLNVVGAM 136
   |||||

```

Search completed: March 13, 2003, 11:28:54
 Job time : 42.9351 secs

